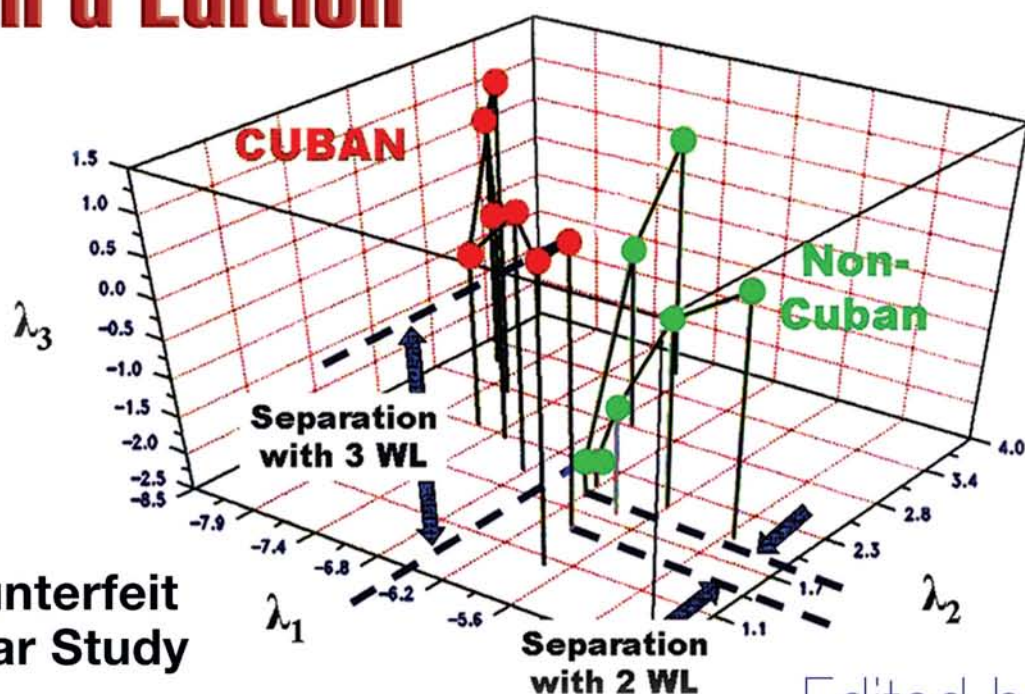


Handbook of Near-Infrared Analysis

Third Edition



Counterfeit
Cigar Study

Edited by

Donald A. Burns
Emil W. Ciurczak

**Handbook of
Near-Infrared
Analysis
Third Edition**

PRACTICAL SPECTROSCOPY A SERIES

1. Infrared and Raman Spectroscopy (in three parts), *edited by Edward G. Brame, Jr., and Jeanette G. Grasselli*
2. X-Ray Spectrometry, *edited by H. K. Herglotz and L. S. Birks*
3. Mass Spectrometry (in two parts), *edited by Charles Merritt, Jr., and Charles N. McEwen*
4. Infrared and Raman Spectroscopy of Polymers, *H. W. Siesler and K. Holland-Moritz*
5. NMR Spectroscopy Techniques, *edited by Cecil Dybowski and Robert L. Lichter*
6. Infrared Microspectroscopy: Theory and Applications, *edited by Robert G. Messerschmidt and Matthew A. Harthcock*
7. Flow Injection Atomic Spectroscopy, *edited by Jose Luis Burguera*
8. Mass Spectrometry of Biological Materials, *edited by Charles N. McEwen and Barbara S. Larsen*
9. Field Desorption Mass Spectrometry, *László Prókai*
10. Chromatography/Fourier Transform Infrared Spectroscopy and Its Applications, *Robert White*
11. Modern NMR Techniques and Their Application in Chemistry, *edited by Alexander I. Popov and Klaas Hallenga*
12. Luminescence Techniques in Chemical and Biochemical Analysis, *edited by Willy R. G. Baeyens, Denis De Keukeleire, and Katherine Korkidis*
13. Handbook of Near-Infrared Analysis, *edited by Donald A. Burns and Emil W. Ciurczak*
14. Handbook of X-Ray Spectrometry: Methods and Techniques, *edited by René E. Van Grieken and Andrzej A. Markowicz*
15. Internal Reflection Spectroscopy: Theory and Applications, *edited by Francis M. Mirabella, Jr.*
16. Microscopic and Spectroscopic Imaging of the Chemical State, *edited by Michael D. Morris*
17. Mathematical Analysis of Spectral Orthogonality, *John H. Kalivas and Patrick M. Lang*
18. Laser Spectroscopy: Techniques and Applications, *E. Roland Menzel*
19. Practical Guide to Infrared Microspectroscopy, *edited by Howard J. Humecki*
20. Quantitative X-ray Spectrometry: Second Edition, *Ron Jenkins, R. W. Gould, and Dale Gedcke*
21. NMR Spectroscopy Techniques: Second Edition, Revised and Expanded, *edited by Martha D. Bruch*
22. Spectrophotometric Reactions, *Irena Nemcova, Ludmila Cermakova, and Jiri Gasparic*
23. Inorganic Mass Spectrometry: Fundamentals and Applications, *edited by Christopher M. Barshick, Douglas C. Duckworth, and David H. Smith*

24. Infrared and Raman Spectroscopy of Biological Materials, *edited by Hans-Ulrich Gremlich and Bing Yan*
25. Near-Infrared Applications In Biotechnology, *edited by Ramesh Raghavachari*
26. Ultrafast Infrared and Raman Spectroscopy, *edited by M. D. Fayer*
27. Handbook of Near-Infrared Analysis: Second Edition, Revised and Expanded, *edited by Donald A. Burns and Emil W. Ciurczak*
28. Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line, *edited by Ian R. Lewis and Howell G. M. Edwards*
29. Handbook of X-Ray Spectrometry: Second Edition, Revised and Expanded, *edited by René E. Van Grieken and Andrzej A. Markowicz*
30. Ultraviolet Spectroscopy and UV Lasers, *edited by Prabhakar Misra and Mark A. Dubinskii*
31. Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy, *Emil W. Ciurczak and James K. Drennen III*
32. Applied Electrospray Mass Spectrometry, *edited by Birendra N. Pramanik, A. K. Ganguly, and Michael L. Gross*
33. Practical Guide to ICP-MS, *edited by Robert Thomas*
34. NMR Spectroscopy of Biological Solids, *edited by A. Ramamoorthy*
35. Handbook of Near-Infrared Analysis, Third Edition, *edited by Donald A. Burns and Emil W. Ciurczak*

Handbook of Near-Infrared Analysis

Third Edition

Edited by
Donald A. Burns
Emil W. Ciurczak



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2008 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-0-8493-7393-0 (Hardcover)

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com>) or contact the Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Handbook of near-infrared analysis / edited by Donald A. Burns and Emil W. Ciurczak. -- 3rd ed.
p. cm. -- (Practical spectroscopy ; 35)

Includes bibliographical references and index.

ISBN 978-0-8493-7393-0

1. Near-infrared spectroscopy. I. Burns, Donald A., 1929- II. Ciurczak, Emil W., 1945- III. Title. IV. Series.

QD96.I5H36 2007

543'.57--dc22

2007000587

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>



We are delighted to dedicate this third edition of the *Handbook of Near-Infrared Analysis* to our good friend and colleague, Karl H. Norris.

Karl is regarded as the “father” of modern near-infrared spectroscopic analysis. He was a major force in the development of the *near-infrared reflection* technology for the simple, accurate, rapid, and inexpensive testing of many quality characteristics of food and grains. This technology has now been widely accepted and has revolutionized the way many chemical analyses are performed. As one can see in the *applications* chapters that follow, the *near-IR* approach to the determination of sample composition now permeates nearly every industry.

After nearly 40 years as an internationally recognized authority in this field, Karl was awarded a well-deserved honorary doctorate by Wilson College on August 10, 2006. This came as a surprise during the 25th anniversary celebration of the *International Diffuse Reflectance Conference* in Chambersburg, PA, before a crowd of nearly 150 colleagues.

Dr. Norris is now retired as leader of the *Instrumentation Research Laboratory* of USDA's *Agricultural Research Service*. He holds several patents, and is the author or co-author of more than 100 research papers.

Way to go, Karl. Keep it up!

Donald A. Burns

Emil W. Ciurczak

Preface

Near-Infrared (NIR) spectroscopy is a technique whose time has arrived. And for good reason — it is unusually fast compared to other analytical techniques (often taking *less than 1 s*), it is nondestructive, and as often as not, no sample preparation is required. It is also remarkably versatile: if samples contain such bonds as C—H, N—H, or O—H, and if the concentration of the analyte exceeds about 0.1% of the total composition, then it is very likely to yield acceptable answers, even in the hands of relatively untrained personnel.

The price to be paid, however, is the preliminary work, which is typical of any chemometric method. The instrument/computer system must be “taught” what is important in the sample. While this task may be time-consuming, it is not difficult. Today’s sophisticated software offers the user such choices of data treatments as multiple linear regression (MLR), partial least squares (PLS), principal components regression (PCR), factor analysis (FA), neural networks (NN), and Fourier transform (FT), among others. The trade-off is a good one: even after several hours (or days) of calibrating, the multiple advantages of analysis by NIR far outweigh the time required for method development.

This book is divided into four parts. Following the Introduction and Background, there is a general section on Instrumentation and Calibration. This is followed by Methods Development, and the depth of NIR’s utility is covered by a broad (if not comprehensive) section on Applications.

The *Handbook of Near-Infrared Analysis* was written for practicing chemists and spectroscopists in analytical, polymer, forage, baking, dairy products, petrochemicals, beverages, pharmaceutical, and textile chemistry who are responsible for methods development or routine analyses where speed, accuracy, and cost are vital factors.

Chapters 2 and 3 have been replaced (Basic Principles and Theories of Diffuse Reflection), several chapters (4,19,25,26,29, and 37) have been updated (Commercial Instrumentation, Analyses of Textiles and of Baked Products, Advances in the Petrochemical Industry, Polymers, Pharmaceutical Applications, and Process Analysis), and some new chapters (12,18,21,28,29,31,32,33,35, and 38) have been added (Process Sensors, Agro-Forestry Systems, Gas Analysis, Use of NIR at the Bowling Alley, PAT in the Pharmaceutical Industry, Nutraceuticals, Detection of Counterfeit Drugs (e.g., Viagra), NIR Photography in Medicine, Biomedical Components in Blood and Serum, and The Detection of Counterfeit Currency and Turquoise).

All this should enable you to assess the potential of NIR for solving problems in your own field (and it could even lead to becoming a hero within your organization). You will discover the relative merits of on-line, in-line, and at-line analyses for process control. You will see how interferences can be removed spectrally rather than physically, and how to extract “hidden” information via derivatives, indicator variables, and other data treatments.

Thanks are due to many people in our lives (i) Linda Burns, who put up with the disappearance of all flat surfaces in our house for those weeks (or was it months?) when piles of papers were everywhere, awaiting some semblance of organization into a book, (ii) the publisher, whose prodding was minimal and whose patience was long, (iii) *the special people in my (Emil’s) life: Alissa, Alex, Adam, Alyssa, and Button (the Wonder Dog)*, and (iv) the many contributors, both original and new. We hope that this general book will be useful to those already using

NIR as well as to those just entering the field and having to make important decisions regarding instrumentation, software, and personnel. Suggestions for improving the next edition are most welcome.

Donald A. Burns

Emil W. Ciurczak

Contributors

M. Kathleen Alam

Sandia National Laboratories
Albuquerque, New Mexico

Carl A. Anderson

Mylan School of Pharmacy
Duquesne University
Pittsburgh, Pennsylvania

Hans-René Bjørsvik

University of Bergen
Bergen, Norway

Karl S. Booksh

Department of Chemistry and
Biochemistry
Arizona State University
Tempe, Arizona

Eric Bouveresse

Department of Analytical Chemistry and
Pharmaceutical Technology
Free University of Brussels
Brussels, Belgium

Marc K. Boysworth

Department of Chemistry and
Biochemistry
Arizona State University
Tempe, Arizona

Chris W. Brown

Department of Chemistry
University of Rhode Island
Kingston, Rhode Island

Bruce Buchanan

S21, LLC
Shenandoah, Virginia

Tom Buist

ENSCO, Inc.
Melbourne, Florida

Donald A. Burns (retired)

Los Alamos National Laboratory
Los Alamos, New Mexico

Douglas S. Burns

ENSCO, Inc.
Melbourne, Florida

Bruce Campbell

Airt Consulting
Douglasville, Georgia

Emil W. Ciurczak

Cadrai Technologies
Golden's Bridge, New York

Donald J. Dahm

Department of Chemistry
Rowan University
Glassboro, New Jersey

Tony Davies

Norwich Near-Infrared
Consultancy
Norwich, United Kingdom

James K. Drennen, III

Mylan School of Pharmacy
Duquesne University
Pittsburgh, Pennsylvania

Yi Ping Du

College of Chemical Engineering
Shandong University of
Technology
Zibo, People's Republic of China

Tom Fearn

Department of Statistical Science
University College London
London, United Kingdom

Rob Frankhuisen

State Institute for Quality Control of Agricultural
Products (RIKILT-DLO)
Wageningen, The Netherlands

Ana Garrido-Varo

Animal Production Department
Faculty of Agriculture and Forestry
Engineering
University of Córdoba
Córdoba, Spain

Subhas Ghosh

College of Technology
Eastern Michigan University
Ypsilanti, Michigan

Peter R. Griffiths

Department of Chemistry
University of Idaho
Moscow, Idaho

Michael J. Hammersley

Wool Research Organisation of
New Zealand, Inc.
Christchurch, New Zealand

Gregory A. Hebner

Sandia National Laboratories
Albuquerque, New Mexico

Peter H. Hindle

Peverel Design Ltd.
Essex, United Kingdom

Ronald Hoogerbrugge

National Institute for Public Health and the
Environment
Bilthoven, The Netherlands

William R. Hruschka

Agricultural Research Service
U.S. Department of Agriculture
Beltsville, Maryland

Jian-Hui Jiang

The State Key Laboratory of Chemo/Bio-Sensing
and Chemometrics
College of Chemistry and Chemical
Engineering
Hunan University
Changsha, People's Republic of China

Tom M. Jones

Research and Development
Sanofi-Aventis
Malvern, Pennsylvania

Sumaporn Kasemsumran

Department of Chemistry
Research Center for Near-Infrared
Spectroscopy
School of Science and Technology
Kwansei-Gakuin University
Sanda, Japan

Gabor J. Kemeny

Kemeny Associates LLC
Middleton, Wisconsin

Lana R. Kington

Bristol-Myers Squibb
Mead Johnson Nutritionals
Evansville, Indiana

Cynthia Kradjel

Integrated Technical Solutions, Inc.
New Rochelle, New York

Kathryn A. Lee

Basking Ridge, New Jersey

Robert A. Lodder

College of Pharmacy
University of Kentucky Medical Center
Lexington, Kentucky

Howard Mark

Mark Electronics
Suffern, New York

Harald Martens

Norwegian University of Science and
Technology
Trondheim, Norway

William J. McCarthy

Thermo Nicolet
Madison, Wisconsin

W. Fred McClure

Department of Biological and Agricultural
Engineering
North Carolina State University
Raleigh, North Carolina

Dennis Mooibroek

National Institute for Public Health and the
Environment
Bilthoven, The Netherlands

Brian G. Osborne

BRI Australia Ltd, North Ryde
New South Wales, Australia

Yukihiro Ozaki

Department of Chemistry
Research Center for Near-Infrared
Spectroscopy
School of Science and Technology
Kwansei-Gakuin University
Sanda, Japan

Emiliano de Pedro

Animal Production Department
Faculty of Agriculture and Forestry
Engineering
University of Córdoba
Córdoba, Spain

Gary E. Ritchie

Scientific Fellow for PAT
United States Pharmacopeia
Rockville, Maryland

James Rodgers

Solutia, Inc.
Gonzalez, Florida

Tor P. Schultz

Forest and Wildlife Research Center
Mississippi State University
Mississippi State, Mississippi

John S. Shenk

College of Agricultural Sciences
The Pennsylvania State University
University Park, Pennsylvania

Alan J. Siegel

ENSCO, Inc.
Melbourne, Florida

Heinz Wilhelm Siesler

Department of Physical Chemistry
University of Duisburg-Essen
Essen, Germany

Suzanne Stanton

Sandia National Laboratories
Albuquerque, New Mexico

Patricia E. Townsend

Wool Research Organisation of New Zealand, Inc.
Christchurch, New Zealand

Aaron A. Urbas

University of Kentucky
Lexington, Kentucky

Marjo J. Vredenburg

National Institute for Public Health and the
Environment
Bilthoven, The Netherlands

Mark O. Westerhaus

Department of Agronomy
The Pennsylvania State University
University Park, Pennsylvania

Phil Williams

Canadian Grain Commission
Winnipeg, Manitoba, Canada

Jerome J. Workman, Jr.

Molecular Spectroscopy and Microanalysis
Thermo Fisher Scientific Inc.
Madison, Wisconsin

Table of Contents

I Introduction and Background	1
Chapter 1	
Historical Development	3
<i>Peter H. Hindle</i>	
Chapter 2	
Basic Principles of Near-Infrared Spectroscopy	7
<i>Heinz W. Siesler</i>	
Chapter 3	
Continuum and Discontinuum Theories of Diffuse Reflection.....	21
<i>Peter R. Griffiths and Donald J. Dahm</i>	
II Instrumentation and Calibration	65
Chapter 4	
Commercial NIR Instrumentation	67
<i>Jerome J. Workman and Donald A. Burns</i>	
Chapter 5	
Fourier Transform Spectrophotometers in the Near-Infrared	79
<i>William J. McCarthy and Gabor J. Kemeny</i>	
Chapter 6	
Analysis Using Fourier Transforms	93
<i>W. Fred McClure</i>	
Chapter 7	
NIR Spectroscopy Calibration Basics	123
<i>Jerome J. Workman, Jr.</i>	
Chapter 8	
Data Analysis: Multilinear Regression and Principal Component Analysis	151
<i>Howard Mark</i>	
Chapter 9	
Data Analysis: Calibration of NIR Instruments by PLS Regression	189
<i>Hans-René Bjørsvik and Harald Martens</i>	

Chapter 10	
Aspects of Multivariate Calibration Applied to Near-Infrared Spectroscopy	207
<i>Marc Kenneth Boysworth and Karl S. Booksh</i>	
Chapter 11	
Transfer of Multivariate Calibration Models Based on Near-Infrared Spectroscopy	231
<i>Eric Bouveresse and Bruce Campbell</i>	
Chapter 12	
Calibration and Validation of Process Sensors	245
<i>Gary E. Ritchie</i>	
III Methods Development	265
Chapter 13	
Sampling, Sample Preparation, and Sample Selection	267
<i>Phil Williams</i>	
Chapter 14	
Indicator Variables: How They May Save Time and Money in NIR Analysis	297
<i>Donald A. Burns</i>	
Chapter 15	
Qualitative Discriminant Analysis	307
<i>Howard Mark</i>	
Chapter 16	
Spectral Reconstruction	333
<i>William R. Hruschka</i>	
IV Applications	345
Chapter 17	
Application of NIR Spectroscopy to Agricultural Products	347
<i>John S. Shenk, Jerome J. Workman, Jr., and Mark O. Westerhaus</i>	
Chapter 18	
The Role of Near-Infrared Spectroscopy in Verifying Label Information in Agro-Forestry Products	387
<i>Ana Garrido-Varo and Emiliano de Pedro</i>	
Chapter 19	
NIR Analysis of Cereal Products	399
<i>B. G. Osborne</i>	
Chapter 20	
NIR Analysis of Dairy Products	415
<i>Rob Frankhuizen</i>	

Chapter 21	
Near-Infrared Spectra of Gases	439
<i>Chris W. Brown</i>	
Chapter 22	
Application for NIR Analysis of Beverages	457
<i>Lana R. Kington and Tom M. Jones</i>	
Chapter 23	
NIR Analysis of Wool	465
<i>Michael J. Hammersley and Patricia E. Townsend</i>	
Chapter 24	
FT/IR vs. NIR: A Study with Lignocellulose	479
<i>Donald A. Burns and Tor P. Schultz</i>	
Chapter 25	
NIR Analysis of Textiles	485
<i>James Rodgers and Subhas Ghosh</i>	
Chapter 26	
Recent Advances in the Use of Near-IR Spectroscopy in the Petrochemical Industry	521
<i>Bruce Buchanan</i>	
Chapter 27	
NIR Analysis of Polymers	529
<i>Cynthia Kradjel and Kathryn A. Lee</i>	
Chapter 28	
Getting NIR Out of the Gutter: The Analysis of Bowling Lane Condition Using Near-Infrared Reflectance Spectroscopy	569
<i>Douglas S. Burns, Alan J. Siegel, and Tom Buist</i>	
Chapter 29	
Process Analytical Technologies (PAT) in the Pharmaceutical Industry	581
<i>Emil W. Ciurczak</i>	
Chapter 30	
Pharmaceutical Applications of Near-Infrared Spectroscopy	585
<i>Carl A. Anderson, James K. Drennen, and Emil W. Ciurczak</i>	
Chapter 31	
NIR in the Dietary Supplement Industry: Qualitative and Quantitative Analysis of Ingredients, Process Blends, and Final Products	613
<i>Cynthia Kradjel</i>	
Chapter 32	
Your Viagras — Genuine, Imitation, or Counterfeit?	631
<i>Marjo J. Vredenbregt, Dennis Mooibroek, and Ronald Hoogerbrugge</i>	

Chapter 33	
Biomedical Applications of Near-Infrared Spectroscopy.....	647
<i>Emil W. Ciurczak</i>	
Chapter 34	
Near-Infrared Spectrometry in Cardiovascular Disease	657
<i>Aaron A. Urbas and Robert A. Lodder</i>	
Chapter 35	
<i>In Vivo</i> and <i>In Vitro</i> Near-Infrared Spectroscopic Determination of Blood Glucose and Other Biomedical Components with Chemometrics.....	673
<i>Yi Ping Du, Sumaporn Kasemsumran, Jian-Hui Jiang, and Yukihiro Ozaki</i>	
Chapter 36	
Plastics Analysis at Two National Laboratories.....	699
Part A	
Resin Identification Using Near-Infrared Spectroscopy and Neural Networks.....	699
<i>M. Kathleen Alam, Suzanne Stanton, and Gregory A. Hebner</i>	
Part B	
Characterization of Plastic and Rubber Waste in a Hot Glovebox	710
<i>Donald A. Burns</i>	
Chapter 37	
Process Analysis.....	717
<i>Gabor J. Kemeny</i>	
Chapter 38	
Detection of Counterfeit Currency and Turquoise	761
<i>Donald A. Burns</i>	
Chapter 39	
Counterfeit Cigars: Can Near Infrared Detect Them?	775
<i>Donald A. Burns</i>	
Chapter 40	
Local Methods and CARNAC-D	781
<i>Tony Davies and Tom Fearn</i>	
Index	797

Part I

Introduction and Background

1 Historical Development

Peter H. Hindle

CONTENTS

1.1 The Discovery of Near-Infrared Radiation	3
1.2 The First Infrared Spectra	4
1.3 A Steady Evolution	4
1.4 The Digital Revolution	5
References	6

1.1 THE DISCOVERY OF NEAR-INFRARED RADIATION

The interaction of light with matter has captured the interest of man over the last two millennia. As early as A.D. 130, Ptolemaeus tabulated the refraction of light for a range of transparent materials, and in 1305, Von Freiburg simulated the structure of the rainbow by using water-filled glass spheres.

By the mid-eighteenth century, through the work of great scientists such as Snell, Huygens, Newton, Bradley, and Priestly, the laws of reflection and refraction of light had been formulated. Both the wave and corpuscular nature of light had been proposed along with measurement of its velocity and adherence to the inverse square law. For the student of the infrared, it is Herschel's discovery of near-infrared (NIR) radiation that is probably of the greatest significance.

Sir William Herschel was a successful musician turned astronomer. Without doubt, he was one of the finest observational astronomers of all time. In 1800, he wrote two papers [1] detailing his study of the heating effect in the spectrum of solar radiation. He used a large glass prism to disperse the sunlight onto three thermometers with carbon-blackened bulbs. Toward the red end of the spectrum, the heating effect became apparent. However, just beyond the red, where there was no visible light, the temperature appeared at its greatest.

Herschel referred to this newly discovered phenomenon as "radiant heat" and the "thermometrical spectrum." Erroneously, he considered this form of energy as being different from light. Whilst his conclusions may appear surprising to us, we must remember that there was no concept of an electromagnetic spectrum, let alone that visible light formed only a small part of it. It was left to Ampere, in 1835, employing the newly invented thermocouple, to demonstrate that NIR had the same optical characteristics as visible light and conclude that they were the same phenomenon. Ampere's contribution, often overlooked, is important because it introduces, for the first time, the concept of the *extended spectrum*.

However, Herschel [2] clearly attached great importance to the analysis of light. In a letter to Professor Patrick Wilson, he concludes:

...And we cannot too minutely enter into an analysis of light, which is the most subtle of all active principles that are concerned with the mechanism of the operation of nature.

By the beginning of the twentieth century, the nature of the electromagnetic spectrum was much better understood. James Clerk Maxwell had formulated his four equations determining the

propagation of light, and the work of Kirchoff, Stefan, and Wien were neatly capped by Max Planck's radiation law in 1900.

Observationally, little progress had been made. Fraunhofer used a newly produced diffraction grating to resolve the sodium "D" lines in a Bunsen-burner flame as early as 1823 and Kirchoff had visually recorded the atomic spectra of many elements by the 1860s. Sadly, lack of suitable detection equipment impeded real progress outside the visible part of the spectrum.

1.2 THE FIRST INFRARED SPECTRA

An important step was taken in the early 1880s. It was noted that the photographic plate, invented in 1829 by Niepce and Daguerre, had some NIR sensitivity. This enabled Abney and Festing [3] to record the spectra of organic liquids in the range 1 to 1.2 μm in 1881. This work was of great significance; not only did it represent the first serious NIR measurements but also the first interpretations, because Abney and Festing recognized both atomic grouping and the importance of the hydrogen bond in the NIR spectrum.

Stimulated by the work of Abney and Festing, W. W. Coblentz [4] constructed a spectrometer using a rock-salt prism and a sensitive thermopile connected to a mirror galvanometer. This instrument was highly susceptible to both vibration and thermal disturbances. After each step in the rotation of the prism, corresponding to each spectral element to be measured, Coblentz had to retire to another room in order to allow the instrument to settle. Each pair of readings (with and without the sample in the beam) was made with the aid of a telescope to observe the galvanometer deflection. It took Coblentz a whole day to obtain a single spectrum. Around 1905 he produced a series of papers and ultimately recorded the spectra of several hundred compounds in the 1- to 15- μm wavelength region.

Coblentz discovered that no two compounds had the same spectrum, even when they had the same complement of elements (e.g., the isomers propan-1-ol and propan-2-ol). Each compound had a unique "fingerprint." However, Coblentz noticed certain patterns in the spectra; for example, all compounds with OH groups, be they alcohols or phenols, absorb in the 2.7 μm region of the spectrum. In this way, many molecular groups were characterized. He also speculated the existence of harmonically related series. Essentially, Coblentz gave chemists a new tool, spectroscopy, where they could obtain some structural information about compounds.

It is interesting to note that contemporaries of Coblentz were working on exciting, new, instrumental designs, which, years later, were to become the mainstay of present-day spectrometry. Rowland developed large, ruled diffraction gratings and concave gratings, in particular, in the 1880s. In 1891, A. A. Michelson [5] published a paper describing the two-beam interferometer.

1.3 A STEADY EVOLUTION

During the first half of the twentieth century many workers extended the spectral database of organic compounds and assigned spectral features to functional groups. While infrared spectroscopy had moved away from being a scientific curiosity it was used very little; suitable spectrometers did not exist and few chemists had access to what instruments there were. Over half a century was to pass between Coblentz's original work and the routine use of spectroscopy as a tool; indeed, two-thirds of a century would pass before routine NIR measurement made its debut.

Possibly, the first quantitative NIR measurement was the determination of atmospheric moisture at the Mount Wilson observatory by F. E. Fowle in 1912 [6] followed, in 1938, by Ellis and Bath [7] who determined amount of water in gelatin. In the early 1940s, Barchewitz [8] analyzed fuels and Barr and Harp [9] published the spectra of some vegetable oils. In the late 1940s Harry Willis, working at ICI, used a prewar spectrometer to characterize polymers and later employed NIR for the measurement of the thickness of polymer films.

The state of the art by the mid-1950s is summarized by Wilbur Kaye [10,11], in 1960 by R. E. Goddu [12] and in 1968 by K. Whetsel [13]. It is interesting to note that up to 1970, only about 50 papers had been written on work concerning NIR.

In the 1930s, lead sulphide (PbS) was being studied as a compound semiconductor, and the advent of the World War II stimulated its development as an infrared detector for heat-sensing purposes. In the 1950s, PbS became available for commercial applications as a very sensitive detector for the 1- to 2.5- μm wavelength region. The NIR region, at last, had a good detector.

Research into NIR spectra (1 to 2.5 μm) as opposed to the mid-infrared (mid-IR) (roughly 2 to 15 μm) had a slow start. Many spectroscopists considered the region too confusing with many weak and overlapping peaks of numerous overtone and combination bands (making assignments difficult). Compared with the mid-IR absorption features were very weak (by two or three orders of magnitude) and, because of the overall complexity, baselines were hard to define. However, two aspects of NIR technology were initially overlooked. First, the PbS detector was very sensitive and because tungsten filament lamps (particularly quartz halogen) were a good source of NIR radiation, diffuse reflection measurements were possible. Second, relatively low-cost instruments could be manufactured because detectors, light sources, and, importantly, optics made from glass were inexpensive.

1.4 THE DIGITAL REVOLUTION

Modern NIR technology relies heavily on the computer (and the microprocessor in particular), not only for its ability to control and acquire data from the instrument, but to facilitate calibration and data analysis. The foundations of data analysis were laid down in the 1930s. Work on the diffuse scattering of light in both transmission and reflection, by Kubelka and Munk [14] in 1931, opened the door to NIR measurements on solids. In 1933, Hotelling [15] wrote a classic paper on principal components analysis (PCA), and Mahalanobis formulated a mathematical approach for representing data clustering and separation in multidimensional space.

In 1938, Alan Turing created the first programmable computer, employing vacuum tubes and relays. By the 1950s, the first commercial computer, UNIVAC, was available. By the mid-1950s, FORTRAN, the first structured, scientific language had been developed by Backus at IBM. The first personal computer (PC) was probably the Altair in 1975, followed in 1977 by the Commodore PET, the same year that saw Bill Gates and Paul Allen found Microsoft. IBM joined the fray with their first PC in 1981 and their designs set the format for compatibility. PC sales of around 300,000 in 1981 rocketed to 3 million in the next year. The PC soon became the driving force behind NIR instrumentation.

Starting in the 1950s there was a growing demand for fast, quantitative determinations of moisture, protein, and oil. Kari Norris, already working for the USDA, was charged with solving the problem for wheat. He then took the bold step of choosing NIR, working with primitive means by today's standards. In 1968, Ben-Gera and Norris published their initial work on applying multiple linear regression (MLR) to the problem of calibration relating to agricultural products. The early 1970s saw the birth of what was to become the laboratory instrument sector of NIR with the emergence of names like Dickey-John, Technicon, and Neotec all in the United States.

At the same time, and quite separately, online process instruments emerged. In Germany, Pier Instrument produced a two-filter sensor with tube-based electronics. In 1970, Infrared Engineering (in the United Kingdom) and Anacon (United States), both employing integrated circuit (IC)-based electronics, entered the marketplace. A few years later Moisture Systems Corporation (United States) entered the field. Online instrumentation is now used for both continuous measurement and process control over a wide range of applications including chemicals, pharmaceuticals, tobacco, food, drinks, and web-based products [16,17].

During the 1980s, the microprocessor was integrated into the designs of most instruments. Much more sophisticated data acquisition and manipulation was now possible. The scope of data treatment display and interpretation was enhanced to include MLR, partial least squares, PCA, and cluster

analysis. Third-party software suppliers have emerged offering a wide choice of data treatments, freeing the user from the constraints of instrument suppliers. Subsequent chapters of this book are devoted in detail to these issues.

NIR technology has evolved rapidly since 1970 and has now gained wide acceptance. In many sectors, it is now the measurement of choice. Its history is far richer than can be presented within the scope of this chapter and so some references for further reading have been included [18–33].

REFERENCES

1. W. Herschel, *Philos. Trans. R. Soc.*, 90: 255–283 (1800).
2. W. Herschel, *William Herschel Chronicles*, vol. 1, William Herschel Museum, Bath, U.K.
3. W. Abney and E. R. Festing, *Philos. Trans. R. Soc.*, 172: 887–918 (1881).
4. W. W. Coblentz, *Investigations of Infrared Spectra Part 1*. Publication No. 35, Carnegie Institute of Washington (1905).
5. A. A. Michelson, *Philos. Mag.*, 31: 256 (1891).
6. F. E. Fowle, *Astrophys. J.*, 35: 149–162 (1912).
7. J. Ellis and J. Bath, *J. Chem. Phys.*, 6: 723 (1938).
8. P. Barchewitz, *J. Chem. Phys.*, 45: 40 (1943).
9. I. Barr and W. Harp, *Phys. Rev.*, 63: 457 (1943).
10. W. Kaye, *Spectrochim. Acta*, 6: 257 (1954).
11. W. Kaye, *Spectrochim. Acta*, 7: 181 (1955).
12. R. E. Goddu, *Adv. Anal. Chem. Inst.*, 1: 347 (1960).
13. K. B. Whetsel, et al., *Anal. Chem.*, 30: 1598 (1958).
14. P. Kubelka and E. Munk, *Zeits. Tech. Physik.*, 12: 593 (1931).
15. H. Hotelling, *J. Ed. Psych.*, 24: 417–441, 489–520 (1933).
16. I. Ben-Gera and K. H. Norris, *J. Feed Sci.*, 64: 33 (1968).
17. I. B. Benson, *Meas. Control*, 22: 45–99 (1989).
18. P. H. Hindle, *Paper Technology and Industry* (Sept. 1984).
19. L. A. Butler, *Cereal Foods World*, 28: 238 (1983).
20. M. S. Day and R. B. Feam, *Lab. Practice*, 31: 328 (1982).
21. G. Birth, *J. Food Sci.*, 44: 949 (1979).
22. S. Borman, *Anal. Chem.*, 56: 934 (1984).
23. D. E. Honigs, *Anal. Instrum.*, 14: 1 (1985).
24. K. H. Norris, *Cereal Foods World*, 31: 578 (1986).
25. D. L. Wetzel, *Anal. Chem.*, 55: 1165 (1983).
26. K. B. Whetsel, *Appl. Spectrosc. Rev.*, 2: 1 (1968).
27. K. Boer, *Trends Anal. Chem.*, 3: 9 (1984).
28. A. Polesello and R. Giangiacomo, *CRC Crit. Rev.*, 18: 203 (1983).
29. C. Stauffer, *Cereal Foods World*, 29: 577 (1984).
30. L. G. Weyer, *Appl. Spectrosc. Rev.*, 21: 1 (1985).
31. R. Moens, *Cereal Foods World*, 25: 518 (1980).
32. E. Stark, et al., *Appl. Spectrosc. Rev.*, 22: 335 (1986).
33. J. Workman, *J. Near Infrared Spectrosc.*, 1: 221–245 (1993).

2 Basic Principles of Near-Infrared Spectroscopy

Heinz W. Siesler

CONTENTS

2.1	Introduction	7
2.2	Basic Principles of Vibrational Spectroscopy	9
2.2.1	The Absorption Techniques of MIR and NIR Spectroscopy	9
2.2.1.1	The Harmonic Oscillator	9
2.2.1.2	The Anharmonic Oscillator	11
2.2.1.3	The Calculation of Overtones and Anharmonicities	12
2.2.1.4	Fermi Resonance, Darling–Dennison Resonance, and the Local Mode Concept	13
2.2.2	The Scattering Technique of Raman Spectroscopy	13
2.2.3	A Comparison of the Qualitative and Quantitative Aspects of Raman, MIR, and NIR Spectroscopy	15
2.2.4	A Comparison of the Basic Instrumentation of Raman, MIR, and NIR Spectroscopy	16
2.2.5	Important Aspects for the Implementation of Raman, MIR, and NIR Spectroscopy in Process Control	18
2.3	Conclusion	18
	References	18

2.1 INTRODUCTION

The increasing demand for product quality improvement and production rationalization in the chemical, petrochemical, polymer, pharmaceutical, cosmetic, food, and agricultural industries has led to the gradual substitution of time-consuming conservative analytical techniques (GC, HPLC, NMR, MS) and nonspecific control procedures (temperature, pressure, pH, dosing weight) by more specific and environmentally compatible analytical tools. In this respect, of the different methods of vibrational (mid-infrared [MIR], near-infrared [NIR] and Raman) spectroscopy, primarily the NIR technique has emerged over the last decade — in combination with light-fiber optics, new in- and on-line probe accessories, and chemometric evaluation procedures — as an extremely powerful tool for industrial quality control and process monitoring.

With this development the wavelength gap between the visible and the MIR region that has over a long period been lying idle is eventually also filled with life and exploited according to its real

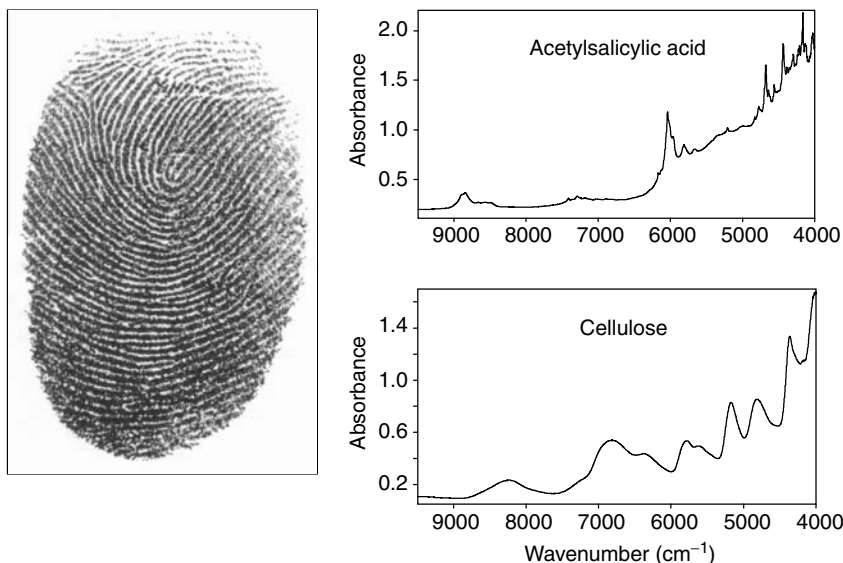


FIGURE 2.1 The NIR spectra of acetylsalicylic acid and cellulose: fingerprints of the chemical composition and the physical state of order.

potential. The period of idling can be mainly contributed to two facts:

- On the one hand, conservative spectroscopists did not accept the frequently broad and overlapped overtone and combination bands of the NIR region as a complementary and useful counterpart to the signals of the fundamental vibrations observed in the Raman and MIR spectra. Figure 2.1 shows the NIR spectra of a crystalline, pharmaceutical active ingredient (acetylsalicylic acid) and an amorphous excipient (cellulose). The characteristic absorption bands with significantly different half-bandwidth clearly demonstrate that this prejudice is not justified and that the NIR spectrum is not only a fingerprint of the chemical composition but also a signature of the physical state of order of the material under investigation.
- Most of the early users of NIR spectroscopy were, with few exceptions, working in the field of agriculture, taking advantage of this new nondestructive analytical tool and having low or no interest to further exploit the spectroscopy behind the data in scientific depth.

Although the situation has not yet changed to the point of equivalent recognition of NIR compared to MIR and Raman spectroscopy, today NIR spectroscopy is at least an accepted technique for industrial applications. It should be emphasized, however, that NIR spectroscopy is not only a routine tool but has also a tremendous research potential, which can provide unique information not accessible by any other technique. Several chapters of this book will certainly prove this statement.

Historically, the discovery of NIR energy is ascribed to W. Herschel in 1800 [1,2]. As far as the development of instrumentation and its breakthrough for industrial applications in the second half of the twentieth century was concerned, it proceeded in technology jumps [3]. In this respect, large credit has to be given to researchers in the field of agricultural science, foremost K. H. Norris [4,5], who have recognized the potential of this technique already in the early fifties. At the same period, with few exceptions [6–8], comparatively low priority had been given to NIR spectroscopy in the chemical industry. This situation is also reflected by the fact that the NIR spectral range was for a long time only offered as a low- or high-wave number add-on to ultraviolet-visible (UV-VIS) or MIR spectrometers, respectively.

This situation has dramatically changed since about the mid-eighties when stand-alone NIR instrumentation became widely available. Since the early nineties, the availability of efficient chemometric evaluation routines, light-fiber optics coupled with specific probes for a multitude of purposes, and the subsequent fast progress in miniaturization, based on new monochromator/detector designs, has launched NIR spectroscopy into a new era for industrial quality and process control.

2.2 BASIC PRINCIPLES OF VIBRATIONAL SPECTROSCOPY

In order to provide a minimum basis to put the different vibrational spectroscopies (Raman, MIR, and NIR) into perspective as far as their theoretical and instrumental fundamentals and their individual advantages are concerned, a short comparative overview is given here. For more detailed information the interested reader is referred to the pertinent literature [9–21].

Although the three techniques are very different in several aspects, their basic physical origin is the same: signals in the MIR, NIR, and Raman spectra of chemical compounds can be observed as a consequence of molecular vibrations. However, while Raman spectroscopy is a scattering technique, MIR and NIR spectroscopy are based on the absorption of radiation (Figure 2.2).

2.2.1 THE ABSORPTION TECHNIQUES OF MIR AND NIR SPECTROSCOPY

2.2.1.1 The Harmonic Oscillator

We will first treat the methods based on the phenomenon of absorption and consider a harmonic diatomic oscillator model where the vibrating masses m_1 and m_2 (Figure 2.2) lead to changes of the internuclear distance $<10\%$. In this case, Hooke's law applies and the potential energy, V , can be

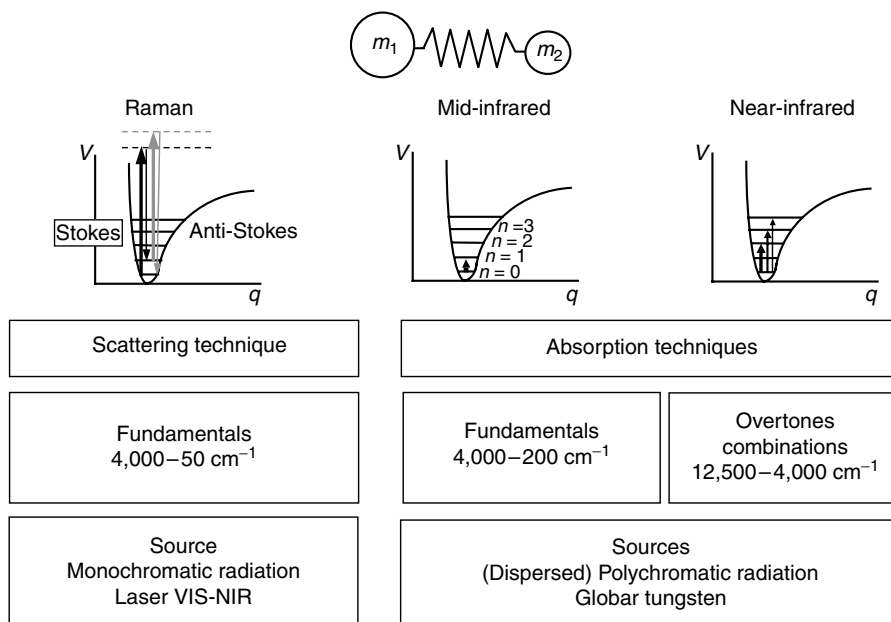


FIGURE 2.2 The principles of Raman, MIR, and NIR spectroscopy (see text).

represented by [22]

$$V = \frac{1}{2}k(r - r_e)^2 = \frac{1}{2}kq^2 \quad (2.1)$$

where k is the force constant of the bond, r is the internuclear distance during the vibration, r_e is the equilibrium internuclear distance, and $q = (r - r_e)$ is the displacement coordinate. The potential energy curve of such an oscillator is parabolic in shape and symmetrical about the equilibrium bond length r_e . This model leads to the vibrational frequency ν_0 :

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (2.2)$$

where the reduced mass m is given by:

$$m = \frac{m_1 m_2}{m_1 + m_2} \quad (2.3)$$

From Equation (2.2) and Equation (2.3), it becomes obvious that the vibrational frequencies are very sensitive to the structure of the investigated compound, and this is the basis for the widespread application of infrared spectroscopy for structure elucidation.

A quantum mechanical treatment by the Schrödinger equation shows that the vibrational energy has only certain discrete values that are given by [22]:

$$E_n = h\nu_0 \left(n + \frac{1}{2} \right) \quad (2.4)$$

where h is Planck's constant, ν_0 is the vibrational frequency defined above and n is the vibrational quantum number that can only have integer values 0, 1, 2, 3, ... and so on. If the energy levels are expressed in wave number units (cm^{-1}), they are given by [3]:

$$G_n = \frac{E_n}{hc} = \bar{\nu}_0 \left(n + \frac{1}{2} \right) \quad (2.5)$$

where c is the speed of light and $\bar{\nu}_0$ is the wave number corresponding to the frequency ν_0

$$\bar{\nu}_0 = \frac{1}{2\pi c} \sqrt{\frac{k}{m}} \quad (2.6)$$

Interaction of infrared radiation with a vibrating molecule, however, is only possible if the electric vector of the radiation oscillates with the same frequency as the molecular dipole moment, μ . Thus, a vibration is infrared active only if the molecular dipole moment is modulated by the vibration [9] and

$$\frac{\partial \mu}{\partial q} \neq 0 \quad (2.7)$$

where q is the vibrational coordinate. The requirement of a dipole moment change during the vibration makes MIR spectroscopy specifically sensitive to polar functionalities (see below and Figure 2.3).

For the harmonic oscillator the energy levels expressed in Equation (2.4) and Equation (2.5) are equidistant and transitions are only allowed between neighboring energy levels with

$$\Delta n = \pm 1 \quad (2.8)$$

Raman	MIR	NIR
$\frac{\partial \alpha}{\partial q} \neq 0$	$\frac{\partial \mu}{\partial q} \neq 0$	$\frac{\partial \mu}{\partial q} \neq 0$ / Anharmonicity $m_2 \ll m_1$
Polarizability	Dipole moment	Anharmonicity
Homonuclear e.g., C=C	Functionalities Polar e.g., C=O	CH/OH/NH Functionalities
High structural selectivity		Low structural selectivity
$I_{\text{Raman}} \propto c$	$\log \frac{I_0}{I} = A = a \cdot b \cdot c$ (Beer's law)	

FIGURE 2.3 Specific characteristics of Raman, MIR, and NIR spectroscopy (I_0 , incident radiation; I , transmitted radiation; A , absorbance; a , absorptivity; b , sample thickness; c , sample concentration) (see text).

According to the Boltzmann distribution, most molecules at room temperature populate the ground level $n = 0$, and consequently the allowed, so-called fundamental, transitions between $n = 0$ and $n = 1$ dominate the vibrational absorption spectrum (Figure 2.2). The potential of MIR spectroscopy as a structure elucidation tool is based on the fact that the majority of absorption bands of chemical compounds corresponding to fundamental vibrations occur in this wave number region (4000 to 200 cm^{-1}).

2.2.1.2 The Anharmonic Oscillator

However, the picture of the harmonic oscillator cannot be retained at larger amplitudes of vibration owing to:

- Repulsive forces between the vibrating atoms.
- The possibility of dissociation when the vibrating bond is strongly extended.

Accordingly, the allowed energy levels for an anharmonic oscillator have to be modified [3,22,23]:

$$G_n = \frac{E_n}{hc} = \bar{\nu}_0 \left(n + \frac{1}{2} \right) - \chi \bar{\nu}_0 \left(n + \frac{1}{2} \right)^2 \quad (2.9)$$

where χ is the anharmonicity constant.

Unlike the harmonic oscillator, energy levels are no longer equidistant and the strict selection rule of Equation (2.8) is expanded to transitions over more than one energy level (see Equation (2.10)). Furthermore, the potential energy curve is represented by an asymmetric Morse function [3,9,22] as shown in Figure 2.2. Generally, a nonlinear molecule containing N atoms will have $3N - 6$ vibrational degrees of freedom, while a linear molecule has only $3N - 5$ [3,24]. The number of vibrational degrees of freedom represents the number of fundamental vibrational frequencies of the molecule or the number of different “normal modes” of vibration. For a given molecule, a normal mode of vibration corresponds to internal atomic motions in which all atoms move in phase with the same frequency, but with different amplitudes. Additionally to these normal vibrations transitions corresponding to

$$\Delta n = \pm 2, \pm 3, \dots \quad (2.10)$$

are now also allowed and are called first, second, and so on, overtones.

Apart from overtones, combinations of different vibrational transitions (sum and difference “tones”) may also be observed. However, the probability of these transitions decreases significantly with their order, and generally the absorption bands corresponding to overtone or combination vibrations have much lower intensity than their fundamental analogs. Contrary to the MIR, the NIR region contains almost exclusively absorption bands that can be assigned to overtone and combination vibrations. Unfortunately, the overlap of these overtone and combination bands strongly decreases the specificity of NIR spectroscopy (especially for interpretation purposes) and was one of the main reasons why this technique has been neglected by conservative spectroscopists for such a long time.

However, the availability of (a) chemometric evaluation procedures for qualitative discrimination and quantitative determination [25–28] and (b) the perception that the low band intensities can be advantageously exploited in terms of larger sample thicknesses and therefore much easier sample handling has eventually led to the breakthrough of the NIR technique.

2.2.1.3 The Calculation of Overtones and Anharmonicities

With Equation (2.9) the wave number position of the fundamental vibration $\bar{\nu}_1$ or an overtone $\bar{\nu}_n$ ($n = 2, 3, \dots$) of the anharmonic oscillator can be given by [23]

$$\bar{\nu}_n = G_n - G_0 = \bar{\nu}_0 n - \chi \bar{\nu}_0 n(n+1) \quad (2.11)$$

$\bar{\nu}_0$ is not directly accessible and from the absorption spectra only the wave numbers $\bar{\nu}_1, \bar{\nu}_2, \dots$ may be obtained. Therefore, we substitute $\bar{\nu}_0$ in Equation (2.9) by

$$\bar{\nu}_0 = \frac{\bar{\nu}_1}{1 - 2\chi} \quad (2.12)$$

and can derive

$$\bar{\nu}_n = \frac{\bar{\nu}_1 n - \bar{\nu}_1 \chi n(n+1)}{1 - 2\chi} \quad (2.13)$$

for $n = 2, 3, 4, \dots$

Thus, if the wave number position $\bar{\nu}_1$ of the fundamental vibration and the anharmonicity constant χ are known, the wave number positions of the overtones can be calculated by Equation (2.13). Alternatively, χ can be calculated if, for example, $\bar{\nu}_1$ and $\bar{\nu}_2$ are known.

The intensities of overtone absorption bands depend on the anharmonicity, and it has been shown [23] that vibrations with low anharmonicity constants also have low overtone intensities. X–H stretching vibrations, for example, have the largest anharmonicity constants and therefore dominate the spectra in the NIR region. Table 2.1 summarizes the anharmonicity constants of the vibrations of some characteristic functionalities.

TABLE 2.1
Anharmonicity Constants χ
for Selected Vibrations [23]

$\chi \nu(\text{CH})$	$\sim 1.9 \times 10^{-2}$
$\chi \nu(\text{CD})$	$\sim 1.5 \times 10^{-2}$
$\chi \nu(\text{CF})$	$\sim 4 \times 10^{-3}$
$\chi \nu(\text{CCl})$	$\sim 6 \times 10^{-3}$
$\chi \nu(\text{C=O})$	$\sim 6.5 \times 10^{-3}$

2.2.1.4 Fermi Resonance, Darling–Dennison Resonance, and the Local Mode Concept

Apart from overtone and combination vibrations, some other characteristic effects may contribute to the appearance of signals in a vibrational spectrum [3].

A resonance that leads to a perturbation of the energy levels can occur if two vibrational levels belong to the same symmetry species and have similar energy. Such an accidental degeneracy of, for example, an overtone or a combination band that has the same symmetry and nearly the same frequency as that of a fundamental vibration is called Fermi resonance [29,30], and this leads to two relatively strong absorption bands that are observed at somewhat higher and lower frequencies than the expected unperturbed frequency positions. When this perturbation takes place, the weaker absorption in the spectrum “steals” intensity from the stronger one. Typical examples of Fermi resonance have been analyzed for the Raman [14] as well as for the NIR spectra of CO₂ [3,31], but this phenomenon has also been reported for numerous other compounds [3,9,14].

A resonance that is of importance in the NIR spectra of water has been discussed by Darling and Dennison [32] but can also occur in other molecules containing symmetrically equivalent X–H bonds. Thus, of the three normal modes of water — ν_2 bending vibration (1595 cm⁻¹), ν_3 anti-symmetric stretching (3756 cm⁻¹), and ν_1 symmetric stretching (3657 cm⁻¹) — the two stretching vibrations absorb at similar wave number positions but belong to different symmetry species and therefore cannot interact directly. However, energy levels of these vibrations associated with specific vibrational quantum numbers n_1 , n_2 , and n_3 [3,32] can interact if they belong to identical symmetry species and have similar energies. These interactions then lead to several pairs of NIR absorption bands with appreciable intensities.

Finally, a few comments shall be made on the concept of local modes as compared to normal modes [3,33–35]. The main idea of the local mode model is to treat a molecule as if it were made up of a set of equivalent diatomic oscillators, and the reason for the local mode behavior at high energy (>8000 cm⁻¹) may be understood qualitatively as follows. As the stretching vibrations are excited to high energy levels, the anharmonicity term $\chi \bar{\nu}_0$ (Equation (2.9)) tends, in certain cases, to overrule the effect of interbond coupling and the vibrations become uncoupled vibrations and occur as “local modes.”

The absorption bands in the spectrum can thus be interpreted as if they originated from an anharmonic diatomic molecule. This is the reason why NIR spectra are often said to become simpler at higher energy. Experimentally, it is found that the inversion from normal to local mode character occurs for high energy transitions corresponding to $\Delta n \geq 3$.

2.2.2 THE SCATTERING TECHNIQUE OF RAMAN SPECTROSCOPY

Whereas scanning MIR and NIR spectrometers operate with a polychromatic source for the individual frequency range (Figure 2.2) from which the sample absorbs specific frequencies corresponding to its molecular vibrational transitions (mostly fundamental vibrations for the MIR and overtone or combination vibrations for the NIR), in Raman spectroscopy the sample is irradiated with monochromatic laser light whose frequency may vary from the VIS to the NIR region. This radiation excites the molecule to a virtual energy state that is far above the vibrational energy levels of this anharmonic oscillator for a VIS-laser and in the range of high overtones for an NIR-laser excitation (Figure 2.2 and Figure 2.4).

From the excited energy level, the molecule may return to the ground state by elastic scattering, thereby emitting the Rayleigh line that has the same frequency as the excitation line and does not contain information in terms of the molecular vibration (this case is not shown in Figure 2.2 and Figure 2.4). If it returns to the first excited vibrational level by inelastic scattering, the emitted Raman line (so-called Stokes line) (Figure 2.2) has a lower frequency (wave number), and the difference to the excitation line corresponds to the energy of the fundamental transition that can also be observed as

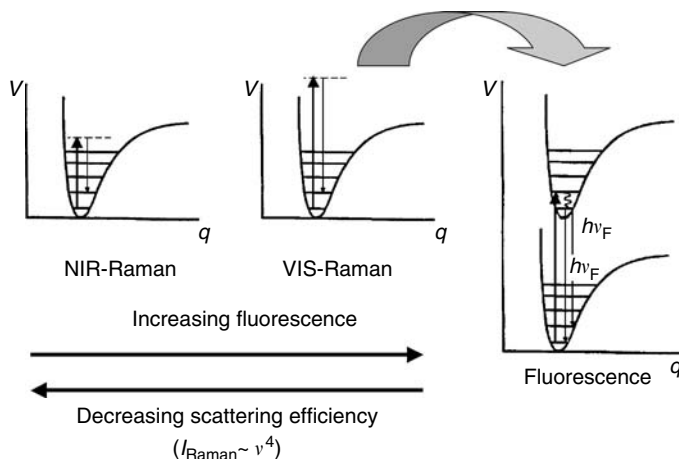


FIGURE 2.4 Fluorescence and scattering efficiency in NIR- and VIS-Raman spectroscopy.

an MIR absorption band. In the case of the anti-Stokes line, where the starting level is the first excited vibrational state and the molecule returns to the ground state by inelastic scattering (Figure 2.2), the emitted Raman line is of higher frequency (here too, the frequency difference to the excitation line corresponds to the fundamental transition) but of lower intensity compared to the Stokes line, due to the lower population of the excited state (law of Boltzmann). Commonly, the Stokes lines are used for practical Raman spectroscopy.

One of the limiting factors for the application of the Raman technique, however, becomes evident by comparing the intensity of the laser source and the scattered radiation [9,15,19–21]

$$I_{\text{Raman}} \approx 10^{-4} I_{\text{Rayleigh}} \approx 10^{-8} I_{\text{source}} \quad (2.14)$$

From these figures it can readily be derived that a sensitive detection of the Raman line alongside an efficient elimination of the Rayleigh line are experimental prerequisites for the successful application of Raman spectroscopy. As shown in Figure 2.2, Raman and MIR spectroscopy cover approximately the same wave number region with the Raman technique extending further into the far-infrared (FIR) region (down to about 50 cm^{-1}) owing to instrumental limitations of the MIR (primarily because of the MIR detector cut-off). In some cases, this additional frequency range is valuable, since it often contains absorptions of lattice modes of molecular crystals that may be very characteristic for a specific polymorph (e.g., of a pharmaceutical active ingredient).

An important relation for the comparison of VIS- vs. NIR-Raman spectroscopy, is the dependence of the scattered Raman intensity I_{Raman} on the fourth power of the excitation frequency ν_{exc}

$$I_{\text{Raman}} \approx \nu_{\text{exc}}^4 \quad (2.15)$$

The impact of this relationship with reference to the application of either VIS- or NIR-Raman spectroscopy for an individual problem will be outlined below.

A similar condition as for MIR spectroscopy holds for the Raman effect and a molecular vibration can only be observed in the Raman spectrum if there is a modulation of the molecular polarizability α [9,15,19–21]:

$$\frac{\partial \alpha}{\partial q} \neq 0 \quad (2.16)$$

Hence, Raman spectroscopy is primarily sensitive to vibrations of homonuclear bonds (Figure 2.3). From the selection rules (Equation (2.7) and Equation (2.16)), it becomes obvious that MIR and

Raman spectroscopy are complementary techniques and the application of both methods can be very helpful for the efficient elucidation of a molecular structure.

2.2.3 A COMPARISON OF THE QUALITATIVE AND QUANTITATIVE ASPECTS OF RAMAN, MIR, AND NIR SPECTROSCOPY

The different excitation conditions of Raman, MIR, and NIR spectroscopy (Figure 2.3) lead to extremely different signal intensities of these techniques for the same vibration of a specific molecular functionality.

NIR spectroscopy covers the wave number range adjacent to the MIR and extends up to the VIS region ($4,000$ to $12,500\text{ cm}^{-1}$) (Figure 2.2). NIR absorptions are based on overtone and combination vibrations of the investigated molecule, and owing to their lower transition probabilities, the intensities usually decrease by a factor of 10 to 100 for each step from the fundamental to the next overtone [3,11,23]. Thus, the intensities of absorption bands successively decrease in the direction from the MIR to the visible region, thereby allowing an adjustment of the sample thickness (from millimeters up to centimeters), depending on the rank of the overtone.

This is a characteristic difference to MIR and Raman spectra, where the signal intensities of the fundamental vibrations vary irregularly over the whole frequency range and depend exclusively on the excitation conditions of the individual molecular vibrations (Equation (2.7) and Equation (2.16)). As pointed out above, these different excitation conditions lead to the complementarity of the Raman and MIR technique as structural elucidation tools, because Raman spectroscopy predominantly focuses on vibrations of homonuclear functionalities (e.g., $\text{C}=\text{C}$, $\text{C}-\text{C}$, $\text{S}-\text{S}$), whereas the most intense MIR absorptions can be traced back to polar groups (e.g., $\text{C}-\text{F}$, $\text{Si}-\text{O}$, $\text{C}=\text{O}$, and $\text{C}-\text{O}-\text{C}$).

NIR spectroscopy, on the other hand, requires — in addition to the dipole moment change — a large mechanical anharmonicity of the vibrating atoms (see Figure 2.3) [3,23]. This becomes evident from the analysis of the NIR spectra of a large variety of compounds, where the overtone and combination bands of CH , OH , and NH functionalities dominate the spectrum, whereas the corresponding overtones of the most intense MIR fundamental absorptions are rarely represented. One reason for this phenomenon is certainly the fact that most of the $\text{X}-\text{H}$ fundamentals absorb at wave numbers $>2000\text{ cm}^{-1}$ so that their first overtones already appear in the NIR frequency range.

The polar groups leading to the most intense fundamental absorptions in the MIR (e.g., $\nu(\text{C}-\text{F})$, $\nu(\text{C}=\text{O})$, $\nu(\text{Si}-\text{O})$) on the other hand absorb at wave numbers $<2000\text{ cm}^{-1}$, so that their first (and sometimes higher) overtones still occur in the MIR region. Owing to the intensity loss for each step from the fundamental to the next overtone, the absorption intensities of these vibrations have become negligible by the time they should occur in the NIR range. The best example in this respect is the $\nu(\text{C}-\text{F})$ absorption band at about 1200 cm^{-1} (e.g., of poly[tetrafluoroethylene]), which is one of the most intense absorption bands in the MIR owing to the large dipole moment of the $\text{C}-\text{F}$ bond.

However, because of the small anharmonicity constant (see Table 2.1), the first and the second overtones that are expected at about 2400 and 3600 cm^{-1} , respectively, have already strongly reduced intensity, and no further overtone vibrations of this functionality can be observed in the NIR region. In fact, poly(tetrafluoroethylene) is used as a nonabsorbing standard material (Spectralon®) for the NIR region.

Anharmonicity plays also an important role in the evaluation of the fundamental and overtone vibration intensities of functionalities with a high hydrogen-bonding tendency such as $\nu(\text{O}-\text{H})$ and $\nu(\text{N}-\text{H})$. Figure 2.5a shows the MIR spectrum of the $\nu(\text{N}-\text{H})$ region of a polyamide 11 (PA 11) film of about $30\text{ }\mu\text{m}$ thickness at room temperature. Under these conditions, the majority of the $\text{N}-\text{H}$ -groups ($\sim 99\%$) occur in the associated, hydrogen-bonded form. This is directly reflected in the very low intensity of the $\nu(\text{N}-\text{H})_{\text{free}}$ absorption at 3450 cm^{-1} relative to the dominating $\nu(\text{N}-\text{H})_{\text{assoc}}$ absorption at 3300 cm^{-1} .

If the same polymer is investigated at room temperature with a film thickness of about $750\text{ }\mu\text{m}$ in the NIR region, the spectrum shown in Figure 2.5b is obtained. Here, the intensity ratio of the

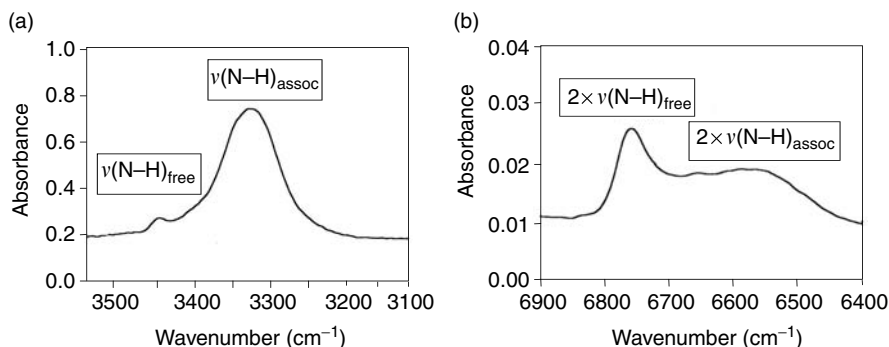


FIGURE 2.5 (a) MIR spectrum of polyamide 11 film in $\nu(\text{N-H})$ fundamental absorption region and (b) NIR spectrum in $2 \times \nu(\text{N-H})$ overtone absorption region.

$2 \times \nu(\text{N-H})_{\text{free}}$ (6760 cm^{-1}) relative to the broad $2 \times \nu(\text{N-H})_{\text{assoc}}$ absorption at about 6510 cm^{-1} is reversed, although the state of order of the polymeric material has not been changed [36]. The explanation for this effect is that, owing to its larger anharmonicity, the intensity of the $\nu(\text{N-H})_{\text{free}}$ overtone absorption is strongly enhanced relative to the corresponding overtone vibration of the associated N-H-groups [37,38]. Hydrogen bonding is equivalent to increasing the mass of the vibrating H-atom, thereby leading to a reduction of mechanical anharmonicity of the $\nu(\text{N-H})_{\text{assoc}}$ vibration and a decrease of its absorption intensity. The uncontrolled use of absorption intensities without proper care for their absorptivities (a) in Beer's law (Figure 2.3) would therefore lead to dramatic errors in the estimation of the extent of hydrogen bonding.

The superposition of many different overtone and combination bands in the NIR region causes a very low structural selectivity for NIR spectra compared to the Raman and MIR analogs where many fundamentals can usually be observed in isolated positions. Nevertheless, NIR spectra should also be assigned in as much detail as possible with reference to their molecular origin [3,36,39]; this allows a more effective application for research purposes and combination with chemometric evaluation procedures. For the assignment of overtones and combination bands in the NIR to their corresponding fundamentals in the MIR, it is recommended that the wave number notation be used instead of the widespread wavelength (nm or μm) scale. It should be mentioned, however, that the wave-number positions of the overtones deviate with increasing multiplicity from the exact multiples of their fundamentals owing to the anharmonicity of the vibrations [3,11,23].

As far as the quantitative evaluation of vibrational spectra is concerned, MIR and NIR spectroscopy follow Beer's law, whereas the Raman intensity I_{Raman} is directly proportional to the concentration of the compound to be determined (Figure 2.3). To avoid compensation problems, in most cases, quantitative Raman spectroscopy is performed with an internal reference signal in the vicinity of the analytical absorption band being analyzed.

An important issue for the implementation of a technique as an industrial routine tool is the sample preparation required for this technique. In this respect, it can be seen from Figure 2.6 that Raman and NIR spectroscopy have considerable advantages over MIR spectroscopy, which usually requires individual sample preparation steps before data acquisition. Only the technique of attenuated total reflection (ATR) [16,40] circumvents time-consuming sampling procedures for MIR spectroscopy.

2.2.4 A COMPARISON OF THE BASIC INSTRUMENTATION OF RAMAN, MIR, AND NIR SPECTROSCOPY

Figure 2.7 summarizes the present state of the most frequently used monochromator/detection principles for the different scanning spectroscopies. As mentioned above, in Raman spectroscopy two

Raman	MIR/ATR	NIR
No sample preparation	Sample preparation required (except ATR)	No sample preparation
Small sample volume (μL) or sample thickness (μm)		Large sample thickness (Up to cm)
Quartz light-fiber optics ($>100\text{ m}$)	Chalcogenide or AgCl light-fiber optics ($<10\text{ m}$)	Quartz light-fiber optics ($>100\text{ m}$)
AT-line/In-line probes	ATR-probes	Transmission transfection diffuse-reflection probes

FIGURE 2.6 Process-control aspects of Raman, MIR, and NIR spectroscopy.

Raman	MIR/ATR	NIR
NIR-Raman (FD) VIS-Raman (CCD)	FT-IR	Grating FT-NIR AOTF Diode-array

FIGURE 2.7 The current monochromator/detection principles of scanning Raman, MIR, and NIR spectrometers.

techniques are presently in current use:

1. Excitation by a VIS-laser (in the range from 400 to 800 nm) combined with monochromatization of the scattered radiation by a holographic grating and simultaneous detection of the dispersed, narrow frequency ranges by a charge-coupled device (CCD) detector.
2. NIR-laser excitation (1064 nm) and measurement in a Fourier-Transform (FT) spectrometer.

In Figure 2.4 the trends of the main limiting factors — fluorescence and low scattering efficiency — have been outlined with reference to the two excitation mechanisms. Thus, both alternatives establish only compromises and the choice of the applied technique depends on the individual problem.

If a molecule is irradiated with visible radiation, it may be excited to an energy level of the next higher electronic state. Return to the ground state or an excited vibrational level of the original electronic state can easily proceed via fluorescence as shown in Figure 2.4. Thus, for a large proportion of samples, irradiation with visible light causes strong fluorescence by additives or impurities (or by the sample itself), which will superimpose and in many cases inundate the Raman spectrum of the sample. The use of NIR-laser excitation confers a number of advantages on a Raman system. Both fluorescence and self-absorption are very much reduced in the Raman signal, and, owing to the lower energy of the excitation radiation, thermal degradation is also less of a problem. However, these advantages are partly neutralized by the disadvantages of using a low-frequency laser as the source.

The NIR-Raman technique is obviously less sensitive due to the ν^4 -dependence of the scattering efficiency (Figure 2.4). Thus, a shift of the excitation line from the VIS region (e.g., Ar⁺-ion laser, 488 nm/20,492 cm⁻¹) to the NIR region (1064 nm/9398 cm⁻¹) reduces the scattering intensity. At 0 cm⁻¹, the sensitivity of a Nd-YAG laser is 23 times lower than that of an Ar laser, and at

$4,000\text{ cm}^{-1}$ this factor has increased to 87 [3,41,42]. As shown in Figure 2.7, however, NIR-Raman spectroscopy is performed on FT-spectrometers, and the sensitivity loss can be compensated by accumulation of multiple scans. As a valuable compromise to suppress fluorescence and at the same time retain an acceptable scattering efficiency, excitation with a diode laser at 785 nm (12739 cm^{-1}) is increasingly used [10,19–21].

As far as MIR spectroscopy is concerned, today almost exclusively FT-based instruments with different interferometer designs are in routine use (Figure 2.7). Contrary to Raman and MIR spectroscopy, scanning NIR spectroscopy offers the largest multiplicity of monochromator/detection principles. Thus, apart from different designs with moving parts, such as grating instruments and FT-spectrometers with Michelson or polarization interferometers (with NIR-transparent quartz wedges) two fast-scanning approaches with no moving parts are available: diode-array systems and acousto-optic tunable filters (AOTF) [10]. Recently, a micro-electro-mechanical-system (MEMS) FT-NIR spectrometer based on a Fabry–Perot interferometer that combines a high spectral resolution and rapid-scanning capability has been brought to the market [43]. Although miniaturization has already progressed significantly with AOTF and diode-array spectrometers, the last mentioned system has launched NIR spectroscopy in a new era of miniaturization and microfabrication technology.

2.2.5 IMPORTANT ASPECTS FOR THE IMPLEMENTATION OF RAMAN, MIR, AND NIR SPECTROSCOPY IN PROCESS CONTROL

In Figure 2.6, the most important aspects for the implementation of the individual spectroscopies as process-monitoring tools are addressed. The very small, representative sample volume or thickness in Raman and MIR/ATR spectroscopy may certainly lead to problems if special care is not taken to avoid the formation of a stationary layer on the reactor window or on the ATR crystal. In this respect, NIR spectroscopy is the method of choice in view of the comparatively large sample volume/thickness involved in these measurements. The ability to separate the spectrometer from the point of sampling is certainly a great advantage for Raman and NIR spectroscopy.

Although light fibers based on chalcogenides, ZrF_4 , and AgCl are also available for MIR spectroscopy, it should be mentioned that their cost, attenuation properties, and mechanical stability are still inferior compared with the well-established quartz fibers. Specific probes are available for all three techniques. NIR spectroscopy offers an especially wide range of in-line, on-line, and at-line transmission and diffuse-reflection probes designed for the measurement of liquids and solids. Large differences can also be identified with respect to the ability of measuring aqueous solutions. Water is an extremely strong absorber in the MIR and also a strong NIR absorber, thereby limiting the available wave number regions in both techniques.

In contrast, it is a weak Raman scatterer and it is recommended to consider Raman spectroscopy as an analytical alternative for aqueous solutions. Care has to be taken, however, with the NIR-Raman FT-technique ($1064\text{ nm}/9398\text{ cm}^{-1}$), because, owing to the absorption of the water-overtone vibration at about 7000 cm^{-1} , the Raman spectrum may be modified relative to the VIS-laser excited Raman spectrum [44].

2.3 CONCLUSION

Over the past years MIR, NIR, and Raman spectroscopy have been further developed to a point where each technique can be considered a potential candidate for industrial quality-control and process-monitoring applications. However, adding up the specific advantages and disadvantages of the individual techniques, NIR spectroscopy is certainly the most flexible and advanced alternative.

REFERENCES

1. W. Herschel, *Philos. Trans. R. Soc. (London)*, 284 (1800).
2. A. M. C. Davies, *NIR News*, **11**, 3 (2000).

3. H. W. Siesler, Y. Ozaki, S. Kawata, and H. M. Heise (eds.), *Near-Infrared Spectroscopy*, Wiley-VCH, Weinheim (2002).
4. A. W. Brant, A. W. Otte, and K. H. Norris, *Food Tech.*, **5**, 356 (1951).
5. G. S. Birth and K. H. Norris, *Food Tech.*, **12**, 592 (1958).
6. W. Kaye, *Spectrochim. Acta*, **6**, 257 (1954).
7. R. G. J. Miller, H. A. Willis, *J. Appl. Chem.*, **6**, 385 (1956).
8. R. F. Goddu and D. A. Delker, *Anal. Chem.*, **32**, 140 (1960).
9. B. Schrader (ed.), *Infrared and Raman Spectroscopy*, Wiley-VCH, Weinheim (1995).
10. J. Coates, *Appl. Spectrosc. Rev.*, **33**(4), 267 (1998).
11. H. W. Siesler, *Makromol. Chem. Macromol. Symp.*, **52**, 113 (1991).
12. J. J. Workman Jr., *Appl. Spectrosc. Rev.*, **34**, 1–89 (1999).
13. P. R. Griffiths and J. A. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley-Interscience, New York (1986).
14. H. Günzler and H. M. Heise, *IR-Spektroskopie*, VCH, Weinheim (1995).
15. P. Hendra, C. Jones, and G. Warnes, *Fourier Transform Raman Spectroscopy*, Ellis Horwood Ltd., Chichester (1991).
16. H. W. Siesler and K. Holland-Moritz, *Infrared and Raman Spectroscopy of Polymers*, Marcel Dekker, New York (1980).
17. J. Chalmers and P. R. Griffiths (eds.), *Handbook of Vibrational Spectroscopy*, John Wiley & Sons Ltd., Chichester (2002).
18. D. A. Burns and E. W. Ciurczak (eds.), *Handbook of Near-Infrared Analysis*, Marcel Dekker, New York (1992).
19. I. R. Lewis and H. G. M. Edwards (eds.), *Handbook of Raman Spectroscopy*, Marcel Dekker, New York (2001).
20. E. Smith and G. Dent, *Modern Raman Spectroscopy — A Practical Approach*, John Wiley & Sons, Chichester (2005).
21. M. J. Pelletier (ed.), *Analytical Applications of Raman Spectroscopy*, Blackwell Publishing, Ames (2000).
22. G. M. Barrow, *Physical Chemistry*, McGraw Hill Inc., New York (1980).
23. W. Groh, *Makromol. Chem.*, **189**, 2861 (1988).
24. E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations — The Theory of Infrared and Raman Vibrational Spectra*, McGraw Hill, New York (1955).
25. D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. de Jong, P. J. Lewi, and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*, Elsevier, Amsterdam, Part A (1997), Part B (1998).
26. H. Martens and T. Naes, *Multivariate Calibration*, Wiley, Chichester (1989).
27. R. Kramer, *Chemometric Techniques for Quantitative Analysis*, Marcel Dekker, New York (1998).
28. T. Naes, T. Isaksson, T. Fearn and T. Davies, *Multivariate Calibration and Classification*, NIR Publications, Chichester (2001).
29. E. Fermi, *Z. Phys.*, **71**, 251 (1931).
30. J. L. Duncan, *Spectrochim. Acta*, **47A**, 1 (1991).
31. M. Buback, J. Schweer, and H. Tups, *Z. Naturforsch.*, **41A**, 505, 512 (1986).
32. B. T. Darling and D. M. Dennison, *Phys. Rev.*, **57**, 128 (1940).
33. R. Mecke, *Z. Phys.*, **81**, 313, 445, 465 (1933).
34. B. R. Henry and W. Siebrand, *J. Chem. Phys.*, **49**, 5369 (1968).
35. I. M. Mills and F. J. Mompean, *Chem. Phys. Lett.*, **124**, 425 (1986).
36. P. Wu and H. W. Siesler, *J. Near Infrared Spectrosc.*, **7**, 65 (1999).
37. S. N. Vinogradov and R. H. Linnell, *Hydrogen Bonding*, Van Nostrand, New York (1971).
38. P. L. Huyskens, W. A. P. Luck, and T. Zeegers-Huyskens (eds.), *Intermolecular Forces*, Springer, Berlin, p. 157 (1991).
39. L. G. Weyer and S.-C. Lo, in *Handbook of Vibrational Spectroscopy* (J. M. Chalmers and P. R. Griffiths (eds.)), John Wiley & Sons, Chichester (2002).
40. F. M. Mirabella Jr. (ed.), *Internal Reflection Spectroscopy*, Marcel Dekker, New York (1993).
41. H. W. Siesler, *Revue de l'Institut Francais du Petrole*, **48**, 223 (1993).
42. A. Hoffmann, PhD-Thesis, Nah-Infrarot Fourier-Transform Raman-Spectroscopic: Genten, Techniken, und Anwendungen University of Essen (1997).
43. www.axsun.com.
44. W.-D. Hergeth, *Chemie Ingenieur Technik*, **70**, 894 (1998).

3 Continuum and Discontinuum Theories of Diffuse Reflection

Peter R. Griffiths and Donald J. Dahm

CONTENTS

3.1	Introduction	22
3.2	Diffuse vs. Directed Radiation	22
3.2.1	Lambert Cosine Law	22
3.3	Absorption and Scatter by a Single Particle	23
3.3.1	Mie Scattering	24
3.4	Continuum Theories of Diffuse Reflection	26
3.4.1	Two-Flux Treatments.....	26
3.4.2	Schuster's Theory	27
3.4.3	Kubelka–Munk Theory.....	28
3.4.4	Discrete Ordinate Approximation.....	34
3.4.5	Diffusion Theory	36
3.4.6	Deviations from the Kubelka–Munk Equation and the Effect of Anisotropic Scatter	37
3.5	Discontinuum Theories of Diffuse Reflection*	41
3.5.1	Theory for an Assembly of Spheres	42
3.5.2	Theory for Sheets and an Assembly Thereof	42
3.5.2.1	The Stokes' Formulas for an Assembly of Sheets	44
3.5.2.2	The Dahm Equation.....	44
3.5.3	The Representative Layer Theory.....	46
3.5.3.1	Model for a Layer Representative of Particulate Solids	46
3.5.3.2	Absorption and Remission of the Representative Layer	46
3.5.3.3	Mathematical Expression of Model	47
3.6	Application of Theory to Model Systems*	48
3.6.1	Example 1: Graphite in NaCl	50
3.6.2	Example 2: Carbazole in a Matrix of Varying Absorption	52
3.6.3	Example 3: Mixture of Wheat and Rape Seed Meal	56
3.7	Experimental Considerations for Reflection Measurements	59
3.7.1	Depth of Penetration.....	59
3.7.2	Effect of Relative Reflectance and Matrix Referencing	61
3.8	Conclusions	62
	Acknowledgment	62
	References	62

3.1 INTRODUCTION

The use of near-infrared (NIR) diffuse reflection for the quantitative analysis of products and commodities is now widely accepted. For many of the algorithms developed to achieve multi-component determinations from the diffuse reflection spectra of powdered samples, a linear dependence of band intensity on analyte concentration is not absolutely mandatory for an analytical result to be obtained. Nonetheless it is probably true to say that all of these algorithms yield the most accurate estimates of concentration when the intensity of each spectral feature is linearly proportional to the analyte concentration. In an analogous manner to transmission spectrometry, reflectance is input to most of these algorithms as $\log(1/R')$, where R' is the reflectance of the sample relative to that of a nonabsorbing standard, such as a ceramic disk. The use of $\log(1/R')$ as the preferred ordinate is contrary to what most physical scientists would consider appropriate for a diffuse reflection measurement on an optically thick sample. Thus an understanding of the theories of diffuse reflection and the validity of the assumptions for each theory should be helpful in understanding the strengths and limitations of NIR diffuse reflection. Perhaps, it will even help to explain why accurate analyses may be made when band intensities are expressed as $\log(1/R')$.

The *continuum* theories of Schuster and of Kubelka and Munk are presented in some detail in this chapter along with a summary of the discrete ordinate approximation of the radiation transfer equation and the diffusion approximation. Together with some of the earlier work on light scattering, an understanding of the importance of the assumptions in arriving at the final solutions that are experimentally valid can be achieved. To a greater or lesser extent, all continuum theories describe model systems and require certain assumptions, such as negligibly small particle size, that are not valid in practice. Several of the drawbacks of these models may be overcome by the use of *discontinuum* models in which the assumptions of a homogenous sample composed of infinitesimally small particles are not invoked. These models are discussed in some detail in the final part of this chapter.

3.2 DIFFUSE VS. DIRECTED RADIATION

The descriptions of diffuse reflection assume that there is a unique direction of light incident upon the sample, and that a plane perpendicular to this direction passes through the sample. A beam of light is called *directed* or *collimated* if all the light is going in a direction perpendicular to the plane of the sample. We then say that the sample is *directly illuminated*. Most commercial spectrometers, to a reasonable approximation, produce a directed incident beam. Light that is scattered from the sample toward a detector on the opposite side of the sample is said to be detected *in transmission*. Light that is scattered from the sample toward a detector on the same side of the sample is said to be detected *in reflection*. If the angle of reflection is equal to the angle of incidence, the reflection is said to be *specular*. Radiation reflected at all other angles is *diffuse*. The sum of the specularly and diffusely reflected radiation is the *remitted* radiation. For samples with a matte finish, especially powdered samples, the specularly reflected radiation is generally of low intensity. Hence measurement of the radiation from this type of sample is frequently known as *diffuse reflection spectrometry*.

3.2.1 LAMBERT COSINE LAW

The phenomenon of diffuse reflection is easily observed in everyday life. Consider for example the intensity of radiation reflected from a completely matte surface such as a sheet of white paper. The remitted radiation is everywhere of the same intensity no matter what the angle of observation or angle of incidence is. It was the same observation that led Lambert [1] to be the first to attempt a mathematical description of diffuse reflection. He proposed that the remitted radiation flux I_r in an area f cm² and solid angle ω steradians (sr) is proportional to the cosine of the angle of incidence α

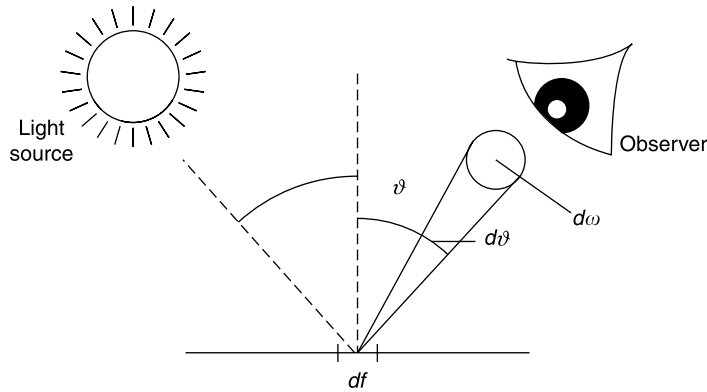


FIGURE 3.1 Schematic representation showing variables used in the Lambert cosine law.

and the angle of observation ϑ (see Figure 3.1), that is,

$$\frac{dI_r/df}{d\omega} = \frac{CS_0}{\pi} \cos \alpha \cos \vartheta = B \cos \vartheta \quad (3.1)$$

where S_0 is the irradiation intensity in W/cm^2 for normal incidence, B is the radiation density or surface brightness in $\text{W}/\text{cm}^2/\text{sr}$, and the constant C is the fraction of the incident radiation flux that is remitted. C is generally less than 1 since some radiation is always absorbed.

Equation (3.1) is known as the Lambert cosine law and, according to Kortüm [2], can be derived from the second law of thermodynamics, although Wendlandt and Hecht [3] disagree. According to Kortüm, it is rigorously valid only for a black-body radiator acting as an ideal diffuse reflector (i.e., the angular distribution of the reflected or remitted radiation is independent of the angle of incidence). It is, however, contradictory to call a black-body radiator an ideal diffuse reflector since all incident radiation is absorbed by a black-body and none is absorbed by an ideal diffuse reflector. An ideal diffuse reflector has never been found in practice and therefore deviations (large and small) always occur from the Lambert cosine law. Various workers, including Wright [4] and Pokrowski [5–8], have reported the results of experimental investigations that were designed to prove or disprove the Lambert cosine law. They found that in general the law holds true only when both the angle of incidence α and the angle of observation ϑ are small.

3.3 ABSORPTION AND SCATTER BY A SINGLE PARTICLE

The terminology used to describe the interaction of light and matter is rather dependent on the size and shape of the particle(s). This is because in limiting situations, some of which we encounter every day, what is observed depends on the nature of the reflecting medium. For example, we see our image in a mirror and call it a *reflection*. In this case, the “particle” has a large flat smooth surface. We look at a glass of milk and say that the white color is a result of the scattering of light by particles or oily droplets in the milk.

When one is describing the interaction of light in situations other than these limiting cases, the result can be somewhat confusing. Terms used by those who study this interaction are described below. The thing that is in common for all the situations is that some of the light may be lost to *absorption*.

When a beam of light is incident on a particle with large flat, smooth surfaces, the word *reflection* describes the process by which light is remitted from the (front) surface. We refer to the passage of light through the particle as *transmission*. If the light entering the particle hits the surface at an

angle, we call the bend in the light path at the surface as *refraction*. A treatment that uses these terms is called *geometric optics*.

When light encounters a very small particle, light is said to be *scattered* by the particle. It might be *back-scattered*, somewhat like reflection, or *scattered in the forward direction*, somewhat like transmission, but here the term *diffuse transmission* refers to the intensity of a *directed* beam of light after it has encountered a collection of particles compared to what it was at the macroscopic surface of the particles. Some of the light is lost to absorption by the particles and some to scatter by the particles, and the rest of the beam is transmitted. The sum of the effects of absorption and scatter is called *extinction*.

What for the case of small particles is called *backward* and *forward scatter* is, for the case of particles with large, flat surfaces, the sum of *reflection* and *transmission*. For infinitesimally small particles, *continuum theories* of diffuse reflection may be applied. As particles get larger, it becomes more likely that the terms for geometrical optics will be applied and *discontinuum theories* are more relevant.

A theory developed by Stokes in the 1860 is generally accepted as describing absorption, remission and transmission by plane parallel layers (sheets). Discussion of this theory is contained in the section on discontinuum theories, along with the discussion of an assembly of sheets. There is no generally accepted theory that describes the absorption and scatter by a collection of spheres as an assembly, although a brief discussion of Melamed's theory is also contained in the discontinuum section. There is, however, a good theoretical description of absorption and scatter by a single, isolated sphere, the cornerstone of which is Mie theory.

3.3.1 MIE SCATTERING

One of the more generally accepted theories of the scattering of light was developed around 1900 by Mie [9]. Mie scattering relates primarily to the scattering of radiation by isolated particles. Only a very brief introduction will be given here, although Kortüm [2] has presented a less abbreviated description. (The reader is referred to References 10–12 for a comprehensive survey, not only of Mie theory but also the theories of Rayleigh, Gans, Born, and others.) Mie obtained solutions to the Maxwell equations [3] that describe the angular distribution of both the intensity and the polarization of scattered radiation for a plane wave scattered once (single scattering) by a particle that can be both dielectric and absorbing. In Mie's description, the particle was spherical, with no limitation on its size. He showed that the angular distribution of scattered radiation for single scattering is not isotropic. The basic equation that was developed by Mie is

$$\frac{I_{\theta s}}{I_0} = \frac{\lambda^2}{8\pi^2 R^2} (i_1 + i_2) \equiv q(\theta_s) \quad (3.2)$$

where $I_{\theta s}$ is the scattered intensity at angle θ at a distance R from the center of the sphere; I_0 is the intensity of the incident radiation, and λ is the wavelength of the incident radiation. The symbols i_1 and i_2 represent complicated functions of the angle of the scattered radiation, the spherical harmonics or their derivatives with respect to the cosine of the angle of scattered radiation, the refractive index of both the sphere and its surrounding medium, and the ratio of the particle circumference to wavelength. Equation (3.2) applies only to the case of a dielectric nonabsorbing particle and unpolarized incident radiation. If the particle is absorbing, the complex refractive index must be used in the determination of i_1 and i_2 .

Mie theory, although general for spherical particles of any size, is valid only for single scattering and therefore directly applicable only to chemical systems in which particles are well separated. For example, scattering by the gases of the atmosphere (the molecules of which are well separated) is a special case of Mie theory, that is, the case where the particle is much smaller than the wavelength of incident radiation. The theory of scattering by particles of this type was developed and explored

by Rayleigh [13,14]. In fact, Kortüm stated that “Rayleigh scattering constitutes a limiting case of the Mie theory for very small particles if only dipole radiation is taken into account” [2].

Mie theory describes the fraction of light that is scattered by a spherical particle in terms of its linear absorption coefficient and the difference between the refractive index of the particle and that of the medium in which it is suspended. It does so uniquely for scatter in every direction. There are some generalizations about the results:

1. For very small particles, the total scatter in the forward direction is equal to the total scatter in the backward direction. (The scatter is mirrored in the forward and backward directions, though not equally at all angles.)
2. For large particles, the scatter in the forward direction is much larger than the scatter in the backward direction.
3. For particles having a size that is a small multiple of the wavelength, there are diffraction ripples in the scattered intensity that can be significant enough to be observed.

There are additional characteristics of radiation that has been scattered from spheres that also have to do with the wave character of light. Let us assume that we have a particle that is large compared to the wavelength of the light. It is then meaningful to talk about reflection from the surface of the sphere, and refraction of the light that enters the sphere. If the absorption of the material of which the particle is composed is large enough, all of the light that enters the sphere is absorbed. The sphere is then *opaque*, and will cast a shadow. However some of the light scattered by the edges of the particle will be diffracted to the center of the shadow of the particle forming diffraction rings. When the incident radiation is a plane wave, this is called a Fraunhofer diffraction pattern, shown in Figure 3.2.

Using the terms of geometric optics, we say that the light that is incident upon the particle is all accounted for by reflection from the external surface of the sphere, the absorption by the sphere and light transmitted through the sphere. The diffraction pattern arises because there is additional light, not directly incident on the sphere, that is scattered by the sphere. In cross section, this light is within a circle around the sphere. The circle has twice the area as the cross section of the sphere, and has a radius of 1.4 times that of the sphere [15]. The area between the cross section of the sphere and the circle is equal to the cross-sectional area of the sphere. The light incident on this area gives rise to the diffraction pattern.

Most applications of NIR reflection analysis involve samples for which it is expected that multiple scattering will take place within the sample. After multiple scattering events, the special effects of scattering from individual spheres tend to be lost. The investigations of Theissing [16] assumed multiple scattering from particles that were sufficiently well separated that interference and phase

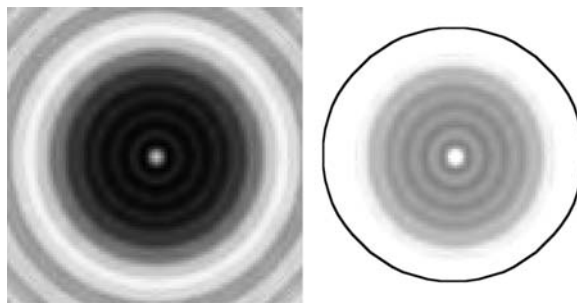


FIGURE 3.2 A Fraunhofer diffraction pattern under two different contrast levels. The circle drawn around the right hand pattern has a radius 1.5 times that of the geometrical shadow of the object. The area between the shadow and the circle represents the area from which the light to form the pattern is obtained. (The patterns were calculated using the Fresnel Diffraction Explorer, which may be obtained from Dager Research: <http://daugerresearch.com/>)

differences between the scattered radiation from the various particles are negligible. *Scattering order* was defined as the number of times a photon is scattered. Theissing found that with an increase in the order of scatter, the forward scattering predicted by Mie theory decreases and the angular distribution of scattered radiation tends to be isotropic. He also found that the larger the ratio of particle circumference to wavelength (designated p), the greater must be the order of scatter to produce an isotropic distribution. For example, if p is 0.6 and the ratio of the refractive index of the sphere to its surrounding medium (designated m) is 1.25, twofold scattering is required for an isotropic distribution of the reflected radiation. But if $p = 5$ and $m = 1.25$, a scattering order of 8 is required for isotropic reflection of radiation. (As a qualitative result, the above is not surprising, given that if a sample is thick enough, almost all the scatter from a collection of spheres is in the backward direction.)

The diameters of particles that are examined by NIR diffuse reflection spectrometry are typically fairly large, on the order of 100 μm , and so p will be large. It is expected that for a sufficiently large number of particles and a sufficiently thick sample (the bounds necessary to define what is sufficient being unknown), multiple scattering does occur for most samples of the type used for NIR reflection analysis. This means that for both already established applications of NIR reflection analysis and potential applications being considered, a theory for multiple scattering within a densely packed medium is required to describe quantitatively the change in reflectance with a change in concentration.

For most samples of the type for which NIR reflection analysis may be possible, the scattering density is large, the ratio of particle circumference to wavelength is much greater than 1, and the particles are so densely packed that the phase relations and interference between scattered beams cannot readily be described. Thus for samples of this type, no general quantitative solution to the problem of multiple scattering has been found. In this case, the scientist must resort to the use of phenomenological theories. (Once the reader has reached the section on discontinuum theories, it will be seen that these theories often have more in common with the theory described for sheets than for spherical particles.) Several of the continuum theories have been rather consistently represented as being a two-flux approximation to the equation of radiative transfer (ERT) (which they are). In the following section, we will occasionally make reference to what they have in common with the theory of sheets.

3.4 CONTINUUM THEORIES OF DIFFUSE REFLECTION

3.4.1 TWO-FLUX TREATMENTS

Much work that has been done on two-flux treatment of the diffuse reflection of radiation has evolved from a general radiation transfer equation. In simple terms, a radiation transfer equation can be written as

$$-dI = \kappa \rho I ds \quad (3.3)$$

An equation such as this describes the change in intensity, dI , of a beam of radiation of a given wavelength in a sample, the density of which is ρ and for which the pathlength is ds . κ corresponds to the attenuation coefficient for the total radiation loss whether that loss is due to scattering or absorption. The general form of the radiation transfer equation that is used in the derivation of most phenomenological theories considers only plane-parallel layers of particles within the sample and can be written as

$$\mu \frac{dI(\tau, \mu)}{d\tau} = -I(\tau, \mu) + \frac{1}{2} \omega_0 \int_{-1}^{+1} p_0(\mu, \mu') I(\tau, \mu') d\mu' \quad (3.4)$$

where μ is the cosine of the angle ϑ with respect to the inward surface normal; μ' is the cosine of the angle ϑ with respect to the outward surface normal; $d\tau$ is the optical thickness and is equal to

$\kappa\rho dx$ where dx is the distance between the boundaries of one plane-parallel layer; I is the intensity of the beam of radiation striking the layer; $\omega_0 = \sigma/(\sigma + \alpha)$ is the albedo* for single scattering, with the scattering and absorption coefficients σ and α , respectively. The scattering phase function $p_0(\mu, \mu')$ denotes the probability for scattering from direction μ' into μ . If every element scatters isotropically, $p_0(\mu, \mu') = 1$ and is independent of the angle between the incident radiation and the scattered radiation. Chandrasekhar has published extensively on radiation transfer and his work is summarized in Reference 17. Other authors who have contributed to the literature on this topic more recently are Truelove [18] and Incropera et al. [19].

3.4.2 SCHUSTER'S THEORY

Schuster was interested in the astrophysical problem of radiation passing through interstellar space. He envisioned a dilute suspension of particles, in which the particles were luminescent, in addition to being absorbing and scattering. The derivation below follows Kortüm [2] in that the luminescence terms are not included. Kortüm calls α (the fraction of light incident upon a particle that is absorbed) the true absorption coefficient of single scattering and σ (the fraction of light incident upon a particle that is scattered) the scattering coefficient for single scattering. It is important to note that this is not what spectroscopists usually mean by a coefficient. In spectroscopy, the coefficient is that quantity, which when multiplied by a pathlength, is used in the calculation of the fraction of incident light that is absorbed or scattered. When the fractions themselves are used as the coefficient (as was done by Kortüm), the unit of thickness of the coefficient is implied to be the particle diameter. Thus coefficients of single scattering are really probabilities.

The classic paper *Radiation Through a Foggy Atmosphere* published in 1905 by Schuster [20] described a particle theory designed to solve a particular problem. However, it was described by Kortüm as an attempt to find a solution of the radiation transfer equation by using the simplified assumption of two oppositely directed radiation fluxes. Radiation traveling in a forward direction through a sample (forward with respect to the direction of the incident radiation) is designated as I . Radiation traveling in the opposite direction is labeled as J . With this simplification, Schuster derived the following two differential equations:

$$\frac{-dI}{d\tau} = (k + s)I - sJ \quad (3.5)$$

$$\frac{dJ}{d\tau} = (k + s)J - sI \quad (3.6)$$

where

$$k = \frac{2\alpha}{\alpha + \sigma} \quad (3.7)$$

and

$$s = \frac{\sigma}{\alpha + \sigma} \quad (3.8)$$

The symbol s used by Schuster is identical to the albedo ω_0 for single scattering. It is relevant to discuss a coefficient of single scattering for a continuum model in that these coefficients relate to the reflectance measured as if the particles in the model were “exploded” apart so that only single scattering could occur.

* In other disciplines, the *albedo* of an object refers to its optical reflectivity, that is, the extent to which it reflects light.

If the boundary conditions are set as

$$I = I_0 \quad \text{at } \tau = 0$$

$$I = I_{(\tau)} \rightarrow 0; \quad J = 0 \quad \text{at } \tau = \tau \quad \text{for } \tau \rightarrow \infty$$

Equations (3.5) and (3.6) are strictly valid only for an *ideally diffusing medium* where there is no change in the degree of dispersion of the light within the sample. The differential equations for the radiant flux in an ideally diffusing medium can be readily integrated. Solutions displayed immediately below may be obtained subject to the boundary conditions that:

1. the intensity of flux I is equal to the incident intensity at a penetration depth of zero, and
2. the intensity of both I and J is zero at “infinite depth,” that is, the sample thickness at which there is no further change in the measured diffuse reflection.

$$R_{\infty} = \frac{J_{(\tau=0)}}{I_0} = \frac{1 - (k/(k + 2s))^{1/2}}{1 + (k/(k + 2s))^{1/2}} \quad (3.9)$$

This equation can be rewritten as

$$\frac{(1 - R_{\infty})^2}{2R_{\infty}} = \frac{k}{s} = \frac{2\alpha}{\sigma} \quad (3.10)$$

Equation (3.10) gives the reflectance behavior for isotropic scattering when two oppositely directed radiation fluxes are assumed in the direction of the surface normal. The function $((1 - R_{\infty})^2/2R_{\infty})$ is commonly known as the Kubelka–Munk (K–M) function and is usually given the symbol $f(R_{\infty})$, although it is interesting to note that in their original paper [21], Kubelka and Munk did not derive this expression. Kubelka actually published the derivation of this equation [22] 17 years later.

We may imagine that the “foggy atmosphere” above is divided into layers, each containing a number of particles. In the limit of infinite dilution, there will be one particle in such a layer. For such a condition, there can be no multiple scatter between particles in the same layer. The above formulation carries with it an implicit assumption that the fraction $(\alpha + \sigma)$ is insignificant compared to 1. This limitation is somewhat relieved by the discontinuous solution

$$f(R_{\infty}) = \frac{(1 - R_{\infty})^2}{2R_{\infty}} = \frac{k}{s} = \frac{\alpha(2 - \alpha - \sigma)}{\sigma} \quad (3.11)$$

This is the discontinuous equivalent of the Schuster equation for isotropic scatter.

3.4.3 KUBELKA–MUNK THEORY

Kubelka and Munk [21] obtained a solution to the radiation transfer problem similar to Schuster’s. While Schuster was considering a “foggy atmosphere,” Kubelka and Munk were considering the “optics of paint layers.” There are significant differences in the way the problem was set up, the most notable being that that Kubelka and Munk solved their differential equations for the case of remission, not isotropic scatter. Their solution resulted in an equation for remission from an infinitely thick coating expressed in terms of the probability of events in an infinitesimal layer of the coating that absorbs (a) and remits (r) a certain constant portion $a \, dx + r \, dx$ of all the light passing through it, where a , the absorption constant and r , the scattering constant, are specific constants of the coating under consideration. You will notice similarity of this approach to that of Dahm and Dahm as shown in Equation (3.83).

In later work, Kubelka [22] published a treatment that is applicable to spectroscopy, which is generally referred to as the K–M theory. It is this later work that we will describe here, following the

treatment of Kortüm [2]. The solution shown here will be obtained through an exponential derivation. A detailed description of a hyperbolic solution can be found in Reference 2.

Unlike Schuster, Kubelka envisioned using the solution for dense systems. While Schuster defined the two constants k and s in terms of the absorption and scattering coefficients for single scattering, Kubelka simply defines K and S in the equations as absorption and scattering coefficients for the densely packed sample as a whole. A tabulation of the variables that are used in their derivation is found in Table 3.1. Figure 3.3 shows a schematic representation of the type of system for which Kubelka and Munk derived their solution.

We will first describe how Kubelka and Munk arrived at the two fundamental differential equations that, once solved, give the simplified solution similar to the one in Equation (3.10). Their stated

TABLE 3.1
Variables Used in the Development of Kubelka's Simplified Solution to the Radiation Transfer Equation

d	Sample layer thickness
$+x$	Downward direction through the sample
$-x$	Upward direction through the sample
$x = 0$	Illuminated surface
$x = d$	Unilluminated surface
I	Radiant flux in $+x$ direction
J	Radiant flux in $-x$ direction
I_0	Incident flux
ϑ	Angle at which a particular ray traverses through dx
dx	An infinitesimal layer
$dx / \cos \vartheta$	Pathlength of a particular ray traversing dx
$d\xi_I$	Average pathlength of radiation passing through dx in the $+x$ direction
$\partial x / \partial \cos \vartheta$	Angular distribution of I 's intensity in the $+x$ direction
$d\xi_J$	Average pathlength of radiation passing through dx in the $-x$ direction
E	Fraction of radiation absorbed per unit pathlength in the sample
σ	Fraction of radiation scattered per unit pathlength in the sample

Note: Neither ε or σ are exactly the same as the corresponding parameters defined by Schuster.

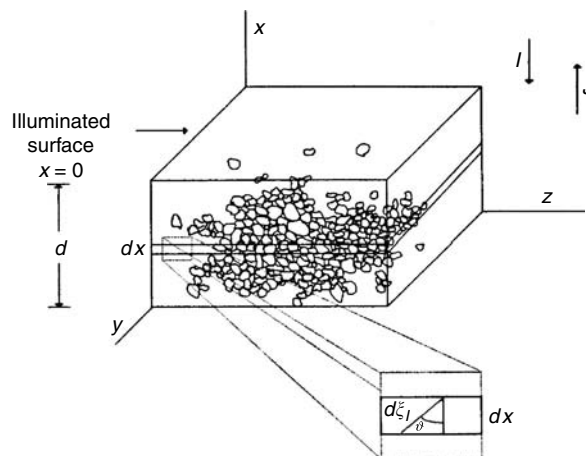


FIGURE 3.3 Schematic representation of a sample for which the K–M equation was derived. Consider the cube as a sample throughout which the particles (only shown in a portion of the sample) are randomly distributed.

assumptions are on page 33. Since they were modeling a dense, particulate system, they needed to take two factors into account. First, because the system was particulate, scattering would diffuse directed radiation. Consequently, in order to avoid the untenable situation that radiation would start out directed and become diffuse during the course of a measurement, they considered systems for which the irradiation began as diffuse. Second, they modeled the particulate system as a sheet, though the word layer is used in Table 3.1. The thickness of the layer is dx , and the continuous mathematics requires that thickness of the layer be infinitesimal. Since the angle through dx , ϑ , that the path of the radiation might follow can be between 0 and 90°, the average pathlength for radiation passing through dx in the $+x$ direction can be found by the following integral:

$$d\xi_I = dx \int_0^{\pi/2} (\partial I / I \partial \vartheta) (d\vartheta / \cos \vartheta) \equiv u dx \quad (3.12)$$

The above formulation holds for a sheet with no voids and for which there are no scattering points within the layer. Kortüm has shown that if the layer were comprised of scattering and absorbing points (i.e., particles of infinitesimal size), the average pathlength through the layer would be twice the geometrical thickness of the layer. Now, if the layer were composed of spherical particles (and an assumption of single scatter within the layer were made) the pathlength through the particles would be unchanged by the angle of incidence on the layer. For real particles we would expect the effect of diffuse radiation to be somewhere between no change and the factor of two increase.

If no absorption or scattering has occurred, the illumination of the layer dx could be described by

$$I d\xi_I = Iu dx \quad (3.13)$$

Since, however, absorption and scattering do occur, the decrease in intensity of the illumination of dx can be shown by including a combination of absorption and scattering coefficients:

$$\begin{aligned} (\varepsilon + \sigma) I d\xi_I &= (\varepsilon + \sigma) Iu dx \\ &= \varepsilon Iu dx + \sigma Iu dx \end{aligned} \quad (3.14)$$

The term $\varepsilon Iu dx$ represents that component of the radiation that is absorbed while the term $\sigma Iu dx$ represents that component of the radiation that is scattered. Radiation traveling in the $-x$ direction (the J flux) also has a corresponding average pathlength through dx

$$d\xi_J = dx \int_0^{\pi/2} (\partial J / J \partial \vartheta) (d\vartheta / \cos \vartheta) \equiv v dx \quad (3.15)$$

Again, if no absorption or scattering were to take place for radiation traveling in this direction, then the illumination of the layer dx in the $-x$ direction would be described by

$$J d\xi_J = Jv dx \quad (3.16)$$

The corresponding equation can be written for the absorption and scattering that occurs for the J radiation flux:

$$(\varepsilon + \sigma) J d\xi_J = (\varepsilon + \sigma) Jv dx = \varepsilon Jv dx + \sigma Jv dx \quad (3.17)$$

Similarly to Equation (3.14), the term $\varepsilon Jv dx$ corresponds to that part of the radiation which is absorbed while the term $\sigma Jv dx$ corresponds to that part of the radiation which is scattered. It is necessary to know, however, the actual change dI or dJ in the radiation fluxes, I and J , which were incident on the layer dx , respectively, after traversing dx . Not only does the absorption and

scattering of the forward radiation flux affect I , but the component of the scattered J radiation flux will also affect I and therefore dI . The same correlation can be made for J and therefore for dJ . It must therefore be true that

$$-dI = \varepsilon u I dx - \sigma v I dx + \sigma u J dx \quad (3.18)$$

$$-dJ = \varepsilon u J dx - \sigma v J dx + \sigma u I dx \quad (3.19)$$

The signs in the above equations are indicative of the fact that as x increases I must decrease and J must increase.

If the sample is an ideal diffuser, then the angular distribution of the radiation flux through a plane layer dx in a given direction is

$$-\frac{\partial I}{\partial \theta} = I \sin 2\theta = 2I \sin \theta \cos \theta \quad (3.20)$$

$$\frac{\partial J}{\partial \theta} = J \sin 2\theta = 2J \sin \theta \cos \theta \quad (3.21)$$

These expressions can then be substituted into Equation (3.11) and Equation (3.14) describing the average pathlength of radiation through dx to solve for this quantity. Doing so we find that $d\xi_I = u = 2$. Likewise, $d\xi_J = v = 2$. If we now substitute for u and v in the differential Equation (3.18) and Equation (3.19) we obtain:

$$-dI = 2\varepsilon I dx - 2\sigma I dx + 2\sigma J dx \quad (3.22)$$

$$dJ = 2\varepsilon J dx - 2\sigma J dx + 2\sigma I dx \quad (3.23)$$

Using Kortüm's notation [2], the absorption coefficient of the material k is equal to ε and the scattering coefficient of the material s is equal to σ . We can then designate $K = 2k$ and $S = 2s$ to obtain the two fundamental simultaneous differential equations from which a simplified solution to the general radiation transfer equation can be found:

$$-dI = K dx - SI dx + SJ dx \quad (3.24)$$

$$dJ = KJ dx - SJ dx + SI dx \quad (3.25)$$

It should be noted here that Kubelka and Munk define scattering differently than does Mie (or Schuster). Mie defines scattering as radiation traveling in any direction after interaction with a particle. Kubelka and Munk defined scattered radiation as only that component of the radiation that is backward reflected into the hemisphere bounded by the plane of the sample's surface. In effect, the defining of S as equal to $2s$ makes S an isotropic scattering coefficient, with scatter equal in both the forward and backward directions.

It may also be noted that, subject to the conditions of infinitesimal particle size, the above differential equations will still hold true even if collimated radiation at an angle of 60° to the surface normal is used instead of diffuse irradiation since for an incident angle of 60° :

$$d\xi_{IJ} = \frac{dx}{\cos 60^\circ} = \frac{1}{0.5} = 2 = u = v \quad (3.26)$$

The differential equations can be simplified by setting

$$\frac{S + K}{S} = 1 + \frac{K}{S} \equiv a \quad (3.27)$$

and $J/I \equiv r$. The simplified form of Equation (3.22) and Equation (3.23) then becomes

$$-dI/S dx = -aI + J \quad (3.28)$$

$$dJ/S dx = -aJ + I \quad (3.29)$$

Dividing the first equation by I and the second equation by J , and then adding the two equations, it is found that

$$dr/S dx = r^2 - 2ar + I \quad (3.30)$$

Using the principle of separation of variables to solve differential equations we obtain

$$I(r^2 - 2ar + 1)^{-1} dr = S dx \quad (3.31)$$

Since the integration must be done over the entire thickness of the sample, the boundaries are

$$x = d: (J/I)_{x=d} = R_g = \text{reflectance of the background} \quad (3.32)$$

$$x = 0: (J/I)_{x=0} = R = \text{reflectance of the sample} \quad (3.33)$$

Equation (3.31) can be integrated using partial fractions where the first step in the fractionation is

$$\frac{dr}{(r^2 - 2ar + 1)} = \frac{1}{(r + (2ar - 1)^{1/2})(r - (2ar - 1)^{1/2})} dr \quad (3.34)$$

The solution of Equation (3.31) can then be found to be

$$\ln \frac{(R' - a - (a^2 - 1)^{1/2})(R_g - a + (a^2 - 1)^{1/2})}{(R_g - a - (a^2 - 1)^{1/2})(R - a + (a^2 - 1)^{1/2})} = 2Sd(a^2 - 1)^{1/2} \quad (3.35)$$

Since it is assumed that the layer is of infinite depth, that is, $d = \infty$ and $R_g = 0$, Equation (3.35) is reduced to

$$(-a - (a^2 - 1)^{1/2})(R_\infty - a(a^2 - 1)^{1/2}) = 0 \quad (3.36)$$

which can be solved for the reflectance R_∞ :

$$R_\infty = \frac{1}{a + (a^2 - 1)^{1/2}} = \frac{1}{1 + K/S + ((K/S)^2 + 2K/S)^{1/2}} \quad (3.37)$$

Rearranging Equations (3.37) in terms of the ratio K/S we obtain an equation similar to the one derived from the work of Schuster:

$$\frac{(1 - 2R_\infty + R_\infty^2)}{2R_\infty} = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{K}{S} \quad (3.38)$$

The above derivation is mathematical and is valid subject to the limitations of the mathematics used. Several assumptions were made, either explicitly or implicitly, in Kubelka's derivation. Most of these should not be viewed as assumptions upon which his solution depended, but rather as a description of the kind of experimental arrangement and samples to which the equations they derived would be

most applicable. The assumptions are listed below:

1. The radiation flux (I and J) travels in two opposite directions.
2. The sample is illuminated with monochromatic radiation of intensity I_0 .
3. The distribution of scattered radiation is isotropic so that all regular (specular) reflection is ignored.
4. The particles in the sample layer (defined as the region between $x = 0$ and $x = d$ are randomly distributed.
5. The particles are very much smaller than the thickness of the sample layer d .
6. The sample layer is subjected only to diffuse irradiation.
7. Particles are much larger than the wavelength of irradiation (so that the scattering coefficient will be independent of wavelength), although if only one wavelength is to be used then this assumption is not relevant.
8. The breadth of the macroscopic sample surface (in the yz plane) is great compared to the depth (d) of the sample and the diameter of the beam of incident radiation (to discriminate against edge effects).
9. The scattering particles are distributed uniformly throughout the entire sample.
10. The absorption by any one particle is small (which is, of course, the case if the particles are infinitesimally small).

A few points might be made about the assumption of isotropic scatter. If there is specular reflection in the scatter, there may be preferential directions of travel through the sample, and the assumption of diffuse radiation will be violated. Assumption 3 points out that the effect of front-surface reflection is ignored in their treatment. This assumption has often been interpreted as meaning that forward and backward scatter from a particle are assumed to be equal. As stated above, related to Equation (3.24) and Equation (3.25), the assumption of isotropic scatter of this kind is also built into their treatment.

Subject to the assumption of infinitesimal particle size, the diffuse reflectance is a function only of the ratio of two constants, K and S , and not of their absolute values. For small particles (i.e., good approximations to infinitesimal particle size), Equation (3.38) can be used to quantitatively determine the concentration. If K is assumed to be proportional to the absorption coefficient obtained in transmission, the equation can be rewritten as shown in Equation (3.39), where a is the absorptivity of the analyte.

$$\frac{(1 - R_\infty)^2}{2R_\infty} = \frac{K}{S} \propto \frac{ac}{S} \quad (3.39)$$

While this is not an exact relationship, as many of the above assumptions are not accurately adhered to in practice, it is very useful for two reasons. First, other treatments of diffuse reflection do not allow R_∞ to be converted into a simple parameter that varies approximately linearly with the concentration of a component of a powdered sample (similarly to absorbance $[\log(1/T)]$ for transmission spectroscopy). If S can be assumed to be constant among a group of samples and the baseline of the spectrum is at a constant value (vide infra), this relationship can be used to quantitatively determine the concentration, c , of an absorbing analyte by preparing a “Beer’s law” plot of the K–M function vs. concentration.

Second, the relationship allows the effect of scattering on diffuse reflection spectra to be understood and forecast. Simply put, the greater the scattering, the weaker the absorption metric. This was demonstrated by Chaffin and Griffiths [23] who measured extended NIR diffuse reflection spectra of three types of polyethylene, a loosely packed powder (high S), an opaque high-density polyethylene (HDPE) bottle (intermediate S) and a translucent milk jug (low S). These spectra are shown in Figure 3.4. The intensity of the first harmonic of the C–H stretching mode at approximately

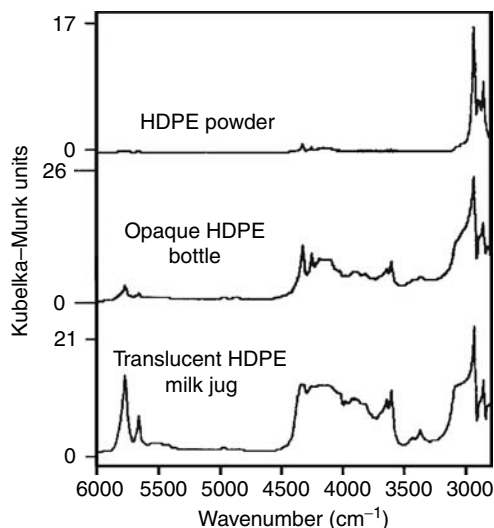


FIGURE 3.4 Extended NIR diffuse reflection spectra of various polyethylene samples. (Reproduced from N. C. Chaffin and P. R. Griffiths, *Appl. Spectrosc.*, 52: 218–221 (1998) by permission of the Society for Applied Spectroscopy Copyright 1998.)

5800 cm^{-1} is greatest for the spectrum of the sample with the lowest scattering coefficient and least for the sample with the highest scattering coefficient.

This result can be explained by a simple consideration of the path of the photons that are back-scattered (remitted) back to the detector. When the scattering coefficient is high, the scattered photons follow a short path in the sample before re-emerging from the face of the sample at which they entered. Thus the pathlength through the absorbing particles is short and the absorption bands are weak. Conversely, when the scattering coefficient is low, the remitted photons that reach the detector pass through many more absorbing molecules in the sample and the absorption bands are relatively strong.

Other scientists in the field [24–28] derived expressions similar to those of Schuster [20] and Kubelka and Munk [21]. Earlier theories developed by Gurevic [29] and Judd [30,31] were shown by Kubelka [22] to be special cases of the K–M theory, while Ingle [32] showed that the formulas derived by Smith [33], Amy [34], and Bruce [35] can be derived from the equations of Kubelka and Munk.

Because the simplified solution obtained by Kubelka is a two-constant equation and therefore experimentally testable, and because so many other workers' derivations are derivable from Kubelka and Munk's work, their solution is the most widely accepted, tested and used. Other workers have derived solutions to the radiation transfer equation that are more complicated than these two-constant formulas. For example, a third constant has been added to account for different fractions of forward and back scattering [36]. Ryde [37,38] included four constants since a difference in the scattering between incident light and internally diffused light is assumed, while Duntley [39] developed a model with eight constants, as a difference between both the absorption and scattering coefficients due to incident and internally diffused radiation was assumed. However, none of these theories is readily applicable in practice, and therefore the treatment of Kubelka is most often applied.

3.4.4 DISCRETE ORDINATE APPROXIMATION

Another two-constant approach, based on a discrete ordinate approximation of the radiation transfer equation [17,40] was recently applied to describe the diffuse reflectance in the NIR [41,42]. In this

approach, the integral in Equation (3.4) is approximated by a weighted sum over discrete directions (i.e., representing the diffuse radiation field inside the sample by radiation fluxes in distinct directions)

$$\int_{-1}^{+1} p_0(\mu, \mu') I(\tau, \mu') d\mu' = \sum_{j=-N}^N a_j p_0(\mu_i, \mu_j) I(\tau, \mu_j) \quad (3.40)$$

where the directions μ_j are chosen to be zeros of Jacobi polynomials

$$P_n^{(0,1)}(2\mu_j - 1) = 0 \quad (3.41)$$

with the corresponding coefficients a_j given by [40]:

$$a_j = \frac{1}{\mu_j^2 (1 - \mu_j) [P_n^{(0,1)'}(2\mu_j - 1)]^2} \quad (3.42)$$

Introducing Equation (3.40) into Equation (3.4) leads to system of differential equations:

$$\mu_i \frac{dI(\tau, \mu_i)}{d\tau} = -I(\tau, \mu_i) + \frac{\omega_0}{2} \sum_{j=-N}^{j=+N} a_j p_0(\mu_i, \mu_j) I(\tau, \mu_j) \quad (3.43)$$

For the case of three radiation fluxes ($j = -1, 0, +1$) the directions and coefficients are given by

$$\mu_{-1} = -\frac{2}{3}, \quad \mu_0 = 0, \quad \mu_{+1} = +\frac{2}{3}, \quad a_{-1} = \frac{3}{4}, \quad a_0 = \frac{1}{2}, \quad a_{+1} = \frac{3}{4} \quad (3.44)$$

and the system of three differential equations can be solved analytically to obtain the diffuse reflectance and transmittance as a function of the optical thickness τ and the albedo ω_0 [43]. With a given sample thickness d and the relations $[\tau = (\alpha + \sigma)d]$ and $[\omega_0 = \sigma/(\alpha + \sigma)]$, the diffuse reflectance and transmittance can also be described as a function of the scattering coefficient σ and the absorption coefficient α (here, the coefficients are defined according to the Mie theory).

In the case of a diffusely illuminated, isotropically scattering and optically thick sample (Kubelka and Munk's assumptions), the three-flux approximation yields the following relation between the ratio α/σ and the diffuse reflectance R_∞ [39]:

$$\frac{\alpha}{\sigma} = \frac{3}{8} \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{3}{8} \frac{K}{S} \quad (3.45)$$

The factor $\frac{3}{8}$ in Equation (3.45) is identical to the result of Mudgett and Richards [44], who related the ratio of the K–M absorption to scattering coefficient to the corresponding ratio of the Mie coefficients. This relationship may be obtained by simple geometrical considerations. The K–M K is an absorption coefficient for the two-flux case. You will recall that the K–M absorption coefficient is equal to twice the absorption coefficient for single scatter or $[K = 2\alpha]$. Then, as similarly written in Equation (3.10):

$$f(R_\infty) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{K}{S} = \frac{2\alpha}{\sigma} \quad (3.46)$$

Mie theory is for spheres of finite size, and the coefficients are for a point of infinitesimal size. For a finite object, the total absorption is related to the volume around this infinitesimal point, and the

scatter is proportional to the cross-sectional area of the object. Including these factors in the above equation, we obtain the two-flux description of absorption and scatter by a finite sphere:

$$f(R_\infty) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{K}{S} = \frac{2\alpha(4/3\pi r^3)}{\sigma(\pi r^2)} = \frac{8\alpha r}{3\sigma} \quad (3.47)$$

Notice that the K–M function for a finite sphere is proportional to the radius of the sphere, and that the coefficients will change with particle size.

Unlike the Mie theory, the discrete ordinate approximation assumes that the sample is continuous matrix of points that absorb and scatter. The values for the coefficients obtained from the discrete ordinate approximation will also change and deviate substantially from the Mie coefficients as the particle size becomes large. This has been called the *hidden mass effect* [42].

To obtain solutions that are relevant for most common spectrometers, for which sample illumination is usually (almost) directional, the intensity $I(\tau, \mu)$ in Equation (3.4) is separated into the reduced incident intensity $I_{\text{dir}}(\tau, \mu)$ and a diffuse intensity $I_{\text{diff}}(\tau, \mu)$

$$I(\tau, \mu) = I_{\text{dir}}(\tau, \mu) + I_{\text{diff}}(\tau, \mu) \quad (3.48)$$

where $I_{\text{dir}}(\tau, \mu)$ is given by

$$I_{\text{dir}}(\tau, \mu) = \frac{F_0}{2\pi} e^{-\tau} \delta(\mu - 1) \quad (3.49)$$

with the directional impinging flux F_0 . Substituting Equation (3.47) and Equation (3.48) into the radiation transfer Equation (3.4) and taking advantage of the δ -function in the presentation of I_{dir} leads to the radiation transfer equation as a function of only the diffuse intensity I_{diff} plus a source function

$$\frac{\omega_0}{4\pi} p(\mu, 1) F_0 e^{-\tau} \quad (3.50)$$

which accounts for the attenuated incident radiation. A more detailed discussion and analytical expressions of the diffuse reflectance and transmittance as a function of τ and ω_0 can be found in the publication by Kuhn et al. [43]. In the case of a directly illuminated, isotropically scattering and optically thick sample, the three-flux approximation yields

$$\frac{\alpha}{\sigma} = \frac{(1 - R_\infty)^2}{2R_\infty} \frac{6}{5(R_\infty + 4)} \quad (3.51)$$

or its inverse

$$R_\infty = \frac{-(4(\alpha/\sigma) + (6/5)) + 2(4(\alpha^2/\sigma^2) + 3(\alpha/\sigma))^{1/2}}{2(\alpha/\sigma) - (6/5)} \quad (3.52)$$

3.4.5 DIFFUSION THEORY

A totally different approach to investigate the radiation transfer is the diffusion approximation, which is often used in biomedical applications [45,46]. The propagation of the photons in a sufficiently thick sample, with a scattering coefficient that is much larger than the absorption coefficient, can be described by a diffusion process (analogous to Fick's law of mass diffusion). Similar to

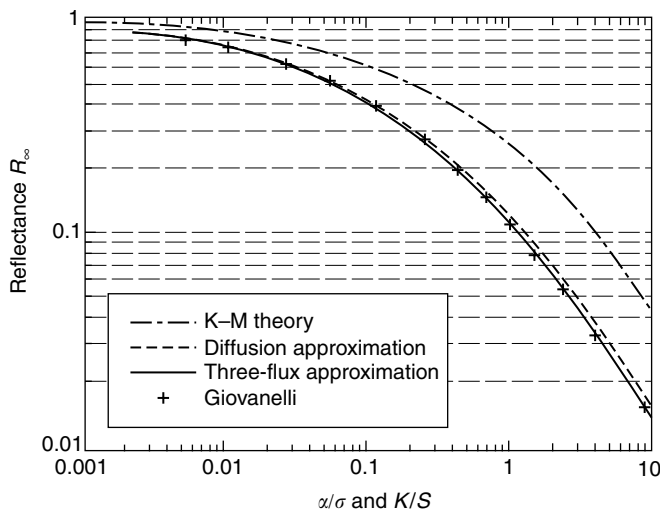


FIGURE 3.5 Diffuse reflectance R_∞ of an isotropic, optically thick sample according to the three-flux approximation, the diffusion approximation, Giovanelli and the K–M theory.

Equation (3.37) and Equation (3.52), the diffuse reflectance of an optically thick sample R_∞ can be expressed as a function of the ratio α/σ , which for directional sample illumination yields [47].

$$R_\infty = \frac{1}{1 + 3(\alpha/\sigma) + (5/3)(3(\alpha^2/\sigma^2) + 3(\alpha/\sigma))^{1/2}} \quad (3.53)$$

The result of applying the different radiative transfer models that describe the diffuse reflectance of an optically thick sample are compared with each other in Figure 3.5. The diffusion approximation (Equation (3.52)), the three-flux approximation (Equation (3.51)) and the data according to a numerical solution of the radiative transfer equation by Giovanelli [48] were calculated for directional sample illumination. The data according to the K–M theory (Equation (3.38)) were obtained for diffuse sample illumination. Thus, the differences between the K–M theory and the other models are due to the different illumination conditions as well as due to differences in the definition of K/S and α/σ . In practice, the three-flux approximation is in very good agreement to the calculations of Giovanelli, whereas the diffusion approximation exhibits increasing deviations for larger α/σ -ratios (corresponding to smaller albedos). For $\alpha/\sigma = 1$ or $\omega = 0.5$, the relative deviations are in the order of 7%.

It should be mentioned that in all the models presented above, specular reflection at the air-sample boundary is neglected; thus these models are only valid for loose powders and have to be used with care for other samples.

3.4.6 DEVIATIONS FROM THE KUBELKA–MUNK EQUATION AND THE EFFECT OF ANISOTROPIC SCATTER

Although the K–M treatment is most often applied to diffuse reflection spectra of dilute dispersions of absorbing materials in a nonabsorbing powdered matrix, it is usually found that, for measurements taken in the mid-infrared or ultraviolet-visible region of the spectrum where the absorptivities of absorption bands may be quite high, plots of $f(R_\infty)$ vs. c , the concentration of the absorbing component, deviate from linearity even at quite low concentrations ($c > \sim 1\%$). For NIR spectra, band absorptivities are usually significantly weaker, so that linear plots of $f(R_\infty)$ vs. c may often be obtained even at fairly high concentrations when the absorptivity of the matrix is zero. On the other

hand, if the analyte has strong absorption bands, NIR diffuse reflection spectra of binary mixtures are no longer adequately described by K–M theory. This was originally thought to be caused by the simple two-flux model (shown in Figure 3.3) breaking down.

The effect of front-surface reflection can be readily seen from the spectra of the two polyethylene bottles in Figure 3.4 [22]. In each case, the K–M function reaches a maximum value of about 18, corresponding to a reflectance, R_∞ , of about 0.05. Since the Fresnel (specular) reflectance at normal incidence for these samples (which have a refractive index of about 1.5) is about 0.04, a reflectance of 0.05 is not unreasonable, especially when the collection optics used for these measurements are taken into account. For NIR spectra, where absorption is so weak that it has little effect of the refractive index, there will always be some maximum value of $f(R_\infty)$ that can be measured in practice (typically about 20). This value will be dependent on the optics used to collect the remitted radiation. One result of reflection from the front surface of the sample is to lead to nonlinear plots of $f(R_\infty)$ against the concentration of an absorbing component in a nonabsorbing matrix, with $f(R_\infty)$ approaching its maximum value asymptotically in a similar way to the effect of stray light in any measurement made with a grating monochromator [49]. Correction for reflection from the front surface of the sample significantly increases the range of linearity.

In a way, because reflection from the first layer of the sample is considered in a different way to the behavior of later layers, this correction can be considered to be an intermediate step between continuous and discontinuous approaches to diffuse reflection. Prior to the development of the discontinuous equations, the nonlinearity was very troublesome to experimenters, and the reason for the departure from linearity was much debated. Besides specular reflection, other proposed culprits were the effect of absorption by the matrix [50,51], and the subject of this section, anisotropic scatter.

One of the assumptions listed at the start of the earlier discussion of K–M theory is that the medium scatters isotropically. Chandrasekhar [17] gave an exact solution of the radiation transfer equation for an isotropically scattering medium, that is, a medium for which the albedo for single scattering, ω_0 , is equal to the scattering phase function, $p(\cos \theta)$, where θ is the angle of scattering from a direction given by $\mu' = \cos \vartheta'$ and ϕ' into a direction given by μ and ϕ . ω_0 is identical to Schuster's scattering coefficient for single scattering, s , given earlier as

$$\omega_0 = \frac{\sigma}{\alpha + \sigma} \quad (3.54)$$

For isotropic scattering, $p(\cos \theta)$ is equal to unity. For anisotropic scattering, the isotropic phase function $p(\cos \theta)$, or ω_0 , is modified by multiplying ω_0 by a factor of $(1 + x \cos \theta)$ so that

$$p(\cos \theta) = \omega_0(1 + x \cos \theta) \quad (3.55)$$

where x is the *anisotropy factor*. Chandrasekhar has also described exact solutions to the radiation transfer equation in this case, but the solution is far too complex to be applicable to the quantitative analysis of powdered mixtures.

The solution has been put into a more tractable form by several workers, including Pitts [52] who developed an approximate solution and Giovanelli [48] who put Pitts's solution into a more useful form (now most commonly known as the Pitts–Giovanelli formula) which gives the reflectance as a function of the direction cosine of the angle of incidence μ_0 as

$$R(\mu_0) = \frac{\omega_0}{2(\chi)^{1/2} + 3 - \omega_0\chi} \left[-x + \frac{3 + (1 - \mu_0)x}{1 + \mu_0(\chi)^{1/2}} \right] \quad (3.56)$$

where

$$\chi = (3 - \omega_0x)(1 - \omega_0) \quad (3.57)$$

If the components of a binary mixture are designated 1 and 2, and the relative weight of those components $W = W_2/W_1$ so that the concentration of component 2 is $W/(W + 1)$, the albedo of single scatter may be expressed as

$$\omega_0 = \frac{1 + P_1 W}{P_2 + P_3 W} \quad (3.58)$$

where $P_1 = \sigma_2/\sigma_1$, $P_2 = (\alpha_1 + \sigma_1)/\sigma_1$, and $P_3 = (\alpha_2 + \sigma_2)/\sigma_1$. Equation (??) may be expressed in terms of the weights, absorption coefficients, and scattering coefficients of the individual components as

$$\omega_0 = \frac{\sigma_1 W_1 + \sigma_2 W_2}{(\alpha_1 + \sigma_1)W_1 + (\alpha_2 + \sigma_2)W_2} \quad (3.59)$$

Thus the Pitts–Giovanelli equation requires four empirical parameters, P_1 , P_2 , P_3 , and x , for a complete characterization of the diffuse reflectance of a binary mixture. Although spectra can be *fit* over a wide concentration range to Equation (3.56) by a suitable selection of the values of these parameters, obtaining good estimates of these parameters a priori is difficult or impossible. Thus the Pitts–Giovanelli treatment is never used in practice to obtain linear plots of a function of the measured reflectance at a wavelength corresponding to an absorption band of the analyte against concentration, even for a binary mixture of an absorbing analyte in a nonabsorbing matrix. This treatment has been shown by Hecht to describe the diffuse reflectance of visible radiation by a model system consisting of soluble absorbers in a liquid medium containing nonabsorbing scattering particles [53,54]. In a related study (again in the visible region of the spectrum), Hecht [55] showed that the Pitts–Giovanelli formula is a good approximation to the equation for radiation transfer even for large values of the anisotropy factor x . He suggested that this result indicates that the breakdown of the K–M theory is not so much due to the failure of the two-flux approximation as to the neglect of anisotropy of scatter.

There is a bit of a paradox in these results. On the one hand, the concept of anisotropy is irrelevant to the two-flux case. The forward scatter is no different than the transmitted beam and so can be neglected. For this reason, the discontinuous methods described in the next section use a concept of remission instead of scatter (as did Kubelka and Munk in their original work). On the other hand, the scatter from a finite particle is undeniably anisotropic. Further, voids may be thought of as producing an extreme case of anisotropic scatter, with all radiation scattered forward, and none back. So while we do not agree that neglect of anisotropy causes breakdown of the two-flux theories, the model of a continuum of anisotropically scattering points is, in principle at least, a better model for real samples than a continuum of isotropically scattering points.

An alternative treatment was developed by Rozenberg [56] and is based on the work of Kuznetsov [57]. Here reflectance is treated as the sum of several components, with each successive term representing increased multiplicity of scatter. The principal independent variable in the Rozenberg treatment is the ratio of the absorption and scattering coefficients β , that is, $\beta = \alpha/\sigma$. If β is sufficiently large, the reflectance R may be described by the equation

$$\frac{1}{R} = \frac{1}{R_0} \frac{(1 + \beta)^2}{1 + \beta/Q} \quad (3.60)$$

where R_0 is the reflectance of the diffusely reflecting matrix in the absence of an absorbing component ($\beta = \beta_0$) and Q is a quantity that defines the relative contribution of higher multiplicities of scatter when $\beta = \beta_0$. β and Q can also be expressed in terms of the reflection r , forward scatter t , and

absorption a of a single layer, where $a + r + t = 1$, as

$$\beta = \frac{a}{r + 1} \quad (3.61)$$

and

$$Q = 1 + \frac{t}{r + t} \quad (3.62)$$

Defining R' as R/R_0 , it can be seen that

$$R' = \frac{1 + \beta/Q}{(1 + \beta)^2} \quad (3.63)$$

In terms of the relative weight W defined above, an expression for β has been derived [58] which is formally analogous to the expression for W_0 in the Pitts–Giovanelli treatment (Equation (3.58)), that is,

$$\beta = \frac{p_1 + p_2 W}{1 + p_3 W} \quad (3.64)$$

where $p_1 = \alpha_1/\sigma_1$, $p_2 = \alpha_2/\sigma_1$ and $p_3 = \sigma_2/\sigma_1$. Again, the visible reflectance of model systems has been fit quite accurately by adjusting the parameters of the Rozenberg equation [58]. In practice, however, neither the Pitts–Giovanelli treatment nor the Rozenberg treatment is particularly relevant for “real-world” samples which often have more than two components and may have a wide range of particle shapes and diameters. Thus a more practical basis for obtaining quantitative data from diffuse reflection spectra must be found.

An alternative way of looking at the nonlinearity of plots of $f(R_\infty)$ vs. c is through discontinuum treatment, which is discussed in more detail later in this chapter. If α is the probability of a photon being absorbed by a layer of the sample, and σ is the probability that it is scattered, the discontinuous equivalent of the K–M equation can be written as

$$f(R_\infty) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{K}{S} = \frac{\alpha(2 - \alpha - \sigma)}{\sigma} \quad (3.65)$$

If we consider infinitesimal layers, the fractions α and σ are very small compared to 2 and can be neglected. In this case, Equation [3.39] will yield a linear plot. However, when finite layers are considered, the function is nonlinear. This will be discussed at length in the second half of this chapter.

Before leaving this topic, one important practical question should be addressed. Since no theory of diffuse reflection, whether continuous as discussed above, or discontinuous, as discussed below, suggests that plots of $\log(1/R_\infty)$ vs. c should be linear, why do almost all practitioners of NIR spectroscopy convert reflectance to $\log(1/R_\infty)$ rather than using $f(R_\infty)$ or any other function that theory suggests should be more relevant? When samples are very carefully prepared, so that the scattering coefficient is constant in all cases, the baseline of the spectrum always has a constant value. In practice, achieving a constant scattering coefficient is exceptionally difficult (In the language of discontinuous mathematics, the distance between the layers can vary significantly depending on the way in which the sample is loaded into the cup). Measurements in the mid-infrared have shown that, provided that the sample is loaded under a constant pressure that is applied for a given amount of time, diffuse reflection spectra can be very reproducible [59,60] and plots of $f(R_\infty)$ vs. c can be linear up to the point at which front surface reflection becomes significant. For NIR diffuse reflection spectrometry, samples are usually simply loaded into a cup pressed against a quartz window. While experimentally simple, this method of presenting the sample to the spectrometer can lead to wide variations in the baseline. Because of the nature of the K–M function,

such baseline variations can significantly affect the value of $f(R_\infty)$ due to an absorption band of the sample [61].

To exemplify this argument, the baseline corrected values of $f(R_\infty)$ and $\log(1/R_\infty)$ for bands absorbing 50% of the baseline energy are listed in the table below for different baseline energies:

Baseline (at A nm) (%)	Maximum absorption (at B nm) (%)	$f(R_\infty)$			$\log(1/R_\infty)$		
		$f(R_\infty)_A$	$f(R_\infty)_B$	$f(R_\infty)_B - f(R_\infty)_A$	$\log(1/R_\infty)_B$	$\log(1/R_\infty)_A$	$\log(1/R_\infty)_A - \log(1/R_\infty)_B$
100	50	0.000	0.250	0.250	0.000	0.301	0.301
80	40	0.025	0.450	0.425	0.097	0.398	0.301
60	30	0.133	0.817	0.684	0.222	0.523	0.301
40	20	0.450	1.600	1.150	0.398	0.699	0.301
20	10	1.6000	4.050	2.450	0.699	1.000	0.301

It can be seen that the baseline-corrected values $[f(R_\infty)_B - f(R_\infty)_A]$ are strongly dependent on the baseline shift, whereas the baseline-corrected values for $[\log(1/R_\infty)_A - \log(1/R_\infty)_B]$ are identical. Even though plots of $\log(1/R_\infty)$ vs. c are nonlinear over wide ranges of concentration, the range of the analyte concentration for many samples that are studied by NIR diffuse reflection is quite small. In this case, the effect of nonlinearity of $\log(1/R_\infty)$ vs. c on the quantitative result is far less than the effect of baseline changes from one sample to the next. In practice, therefore, most practitioners convert their data to $\log(1/R_\infty)$ rather than $f(R_\infty)$.

3.5 DISCONTINUUM THEORIES OF DIFFUSE REFLECTION*

A continuum theory implicitly assumes a model for the absorption by and scatter from a particle of infinitesimal size. This model is only a reasonable approximation for samples in which the fraction of light absorbed by an individual particle is a very small fraction of the light incident upon it. The advantage of this model is that it is simple, though the mathematics that describe it are not. The discontinuum method we will describe uses mathematics no more complex than the continuum theories. However, the description of the sample is more complex. This is both the power and limitation of the discontinuum theories: they can describe more complex situations, but doing so requires a more detailed description.

Currently, no continuum or discontinuum theory produces a function that is linear with concentration of absorbers over large concentration ranges. However, the discontinuum theory can provide an understanding as to why the functions behave the way they do. A discontinuum theory can make visible that which is hidden in the implicit assumptions of continuous mathematics. The largest advantage of the discontinuous theory to a spectroscopist is that models appropriate for mixtures can be used to describe a sample.

As noted above, the phenomenological two-flux theories that have been developed on the basis of the radiation transfer equation can be considered *continuum* theories. Continuum theories consider the absorption and scattering coefficients as properties of an irradiated isotropic layer of infinitesimal thickness. On the other hand, *discontinuum* theories consider layers containing a collection of particles. Consequently, the thickness of a layer is dictated by the size of the scattering and absorbing particles. Optical constants can then be determined from the scattering and absorption properties of these particles.

* Sections 3.5 and 3.6 are drawn in large part from a book: *Interpreting Diffuse Reflectance and Transmittance: A Theoretical Introduction to Absorption Spectroscopy of Scattering Materials* by Donald J. Dahm and Kevin D. Dahm, NIR Publications (2007).

Recently, the discontinuous approach has been applied by Dahm and Dahm [62,63] to the two-flux results obtained from the radiation transfer equation. This has resulted in being able to recast the results obtained from the continuous approach in terms of a layer of particles. This has been termed the *representative layer theory*. In the discussion of deviations from the K–M equation, we gave without proof an analogous mathematical expression reached by the discontinuous treatment (Equation (3.65)). In this section we give the derivation of some of the more important formulas in the discontinuous treatment.

3.5.1 THEORY FOR AN ASSEMBLY OF SPHERES

Melamed developed an elegant mathematical description of absorption and scatter from an assembly of close packed spheres [64]. Unfortunately, it is remembered more for its failings than its elegance. The Melamed model embodied two assumptions: There would be reflection from the front external surface of a particle (that would follow Lambert's cosine law for a sphere), and there would be isotropic scatter from inside the particle. Within the sphere, the model took into account an infinite number of internal reflections. The model predicts that reflectance would reach a maximum as the relative refractive index approached zero. Of course, if there is no refractive index difference between that of the particle and the medium, there is no reflection from the particle.

While the primary failings of the theory were “fixed” with a modification by Simmons [65], it was Simmons himself who pointed out that simple scattering models seem to be as useful as more complex ones. Recent work in discontinuous theories has been dominated by using a two-flux model and the mathematics of plane parallel layers (sheets).

3.5.2 THEORY FOR SHEETS AND AN ASSEMBLY THEREOF

Consider a sheet under directed illumination. A sheet is an object with two large, flat, smooth, parallel surfaces with two dimensions much larger than the third. This third dimension will be referred to as the thickness. The layers are illuminated with a directed beam from a direction perpendicular to the large dimensions. The diameter of the beam is small compared to the front surface area of the layer that it strikes.

At normal incidence, the reflectance from a surface, r_0 , may be calculated from the index of refraction of the material, n_1 , and n_0 , the index of refraction of the dispersing medium (in this case air) using the Fresnel equation

$$r_0 = \frac{(n_1 - n_0)^2}{(n_1 + n_0)^2} \quad (3.66)$$

This formula neglects the effects of absorption on remission. Because in NIR, the band absorptivities are very small, the refractive index, and hence the reflectance, of a surface should not vary significantly with absorption. This is not the case in the mid-infrared region. In this case, not only the effect of absorption on r_0 , but also the fact that the refractive index varies significantly across a strong absorption band, an effect known as *anomalous dispersion*, must be taken into account when applying the Fresnel equations.

Within a layer, there is a transmission loss due to absorption, which can be calculated from the Bouguer–Lambert law by $[1 - \exp(-kd)]$, where k is the linear absorption coefficient of the material making up the layer and d is the thickness of the layer.[†] Consequently, the light incident

[†] The term Bouguer–Lambert law is not familiar to many spectroscopists. The term Beer–Lambert law or merely Beer's law is frequently used in its place. Technically, Beer's law refers to the observation that the contribution of an absorber to the absorbance of a sample is proportional to the concentration of the absorber. The symbol k is referred to by some spectroscopists as the *Beer–Lambert absorption coefficient*. Because of the possibility of decadic or napierian absorbance and the various units by which concentration can be expressed, several different quantities are all Beer–Lambert absorption coefficients. The term absorptivity is commonly used in equations for decadic absorbance and can include concentration in any units. The term *linear absorption coefficient* is the usual name for the *linear napierian absorption coefficient* of a pure material.

on the back surface has been diminished by absorption, and the reflection from it is correspondingly diminished.

The fraction of incident light remitted by a single layer, R_1 , the total transmittance through a layer, T_1 , and the fraction of light absorbed by a layer, A_1 , can be calculated from the Stokes equations [3,66]:

$$R_1 = r_0 + \frac{(1 - r_0)^2 r_0 \exp(-2kd)}{1 - r_0^2 \exp(-2kd)} \quad (3.67)$$

$$T_1 = \frac{(1 - r_0)^2 \exp(-kd)}{1 - r_0^2 \exp(-2kd)} \quad (3.68)$$

$$A_1 = 1 - R_1 - T_1 \quad (3.69)$$

The expression for R_1 above reveals an important physical insight. The expression $\exp(-2kd)$ is always between 0 and 1 and appears in a negative term of the denominator. Thus, for the case of a constant r_0 , the effect of increasing absorption will be to reduce the remission of radiation that has penetrated into a single layer. Consequently, even if the remission from a *single surface* is independent of absorption, remission from an absorbing *layer or particle*, which has both a front and rear surface, does depend on absorption. The assumption of a constant r_0 , which depends on a constant refractive index, is a good one in the NIR region. In the mid-infrared, the case is more complicated because the refractive index changes across strong absorption bands, thereby increasing r_0 . In such a case it is even more likely that the remission from an absorbing *layer or particle* depends on absorption.

Derivation of these formulas involves assuming that a fraction r_0 is reflected by the front surface and a fraction $(1 - r_0)$ is transmitted into the particle, where it is attenuated, that is, a fraction $[(1 - r_0) e^{-kd}]$ reaches the back surface. Here a fraction $[(1 - r_0)^2 e^{-kd}]$ leaves the particle in the forward direction contributing to its transmission. A fraction $[r_0(1 - r_0) e^{-kd}]$ returns back toward the front surface and is again attenuated by absorption, with $[r_0(1 - r_0) e^{-2kd}]$ reaching the front surface. Here a fraction $[r_0(1 - r_0)^2 e^{-2kd}]$ leaves the particle and contributes to the reflectance. This process continues indefinitely and is described as infinite series that converge to the above expressions.

$$R_1 = r_0 + r_0(1 - r_0)^2 e^{-2kd} + r_0^3(1 - r_0)^2 e^{-4kd} + \dots + r_0^{2n-1}(1 - r_0)^2 e^{-2nkd} \quad (3.70)$$

$$T_1 = (1 - r_0)^2 e^{-kd} + r_0^2(1 - r_0)^2 e^{-3kd} + \dots + r_0^{2n-1}(1 - r_0)^2 e^{-(2n-1)kd} \quad (3.71)$$

It is assumed that there is no divergence of the beam, and that on every pass through the layer, the light will travel exactly the distance d . This means that, with conventional spectroscopic equipment, there would be no remission detected from the sample, because it would all be reflected back directly into the incident beam.

The above derivation may be generalized for the case where r_e is the fraction of light reflected from the surface as the light *enters* the layer and r_f is the fraction of light reflected at the surface when the light *leaves* the layer. Note that the layer is still considered symmetrical in this model: the front and back surfaces both behave the same way, the distinction between r_e and r_f concerns whether light is traveling from the surrounding medium to the layer, or from the layer to the surrounding medium. For this situation, the above equations become:

$$R_1 = r_e + \frac{(1 - r_e)(1 - r_f)r_f \exp(-2kd)}{1 - r_f^2 \exp(-2kd)} \quad (3.72)$$

$$T_1 = \frac{(1 - r_e)(1 - r_f) \exp(-kd)}{1 - r_e^2 \exp(-2kd)} \quad (3.73)$$

$$A_1 = 1 - R_1 - T_1 \quad (3.74)$$

For the case of diffuse illumination, there is a difference in the distance traveled for diffuse light, as opposed to directed light. We may define a new absorption coefficient, K , which is dependent on the actual distance that the light travels through the sheet. This is different from the distance traveled by directed radiation, where the distance traveled is equal to the sample thickness. For diffuse illumination of plane parallel particles, the relationship between the two absorption coefficients is

$$\exp(-Kd) = \int_0^{\pi/2} \frac{\exp(-kd)}{\cos(\theta)} d\theta \quad (3.75)$$

The quantities $\exp(-kd)$ and $\exp(-Kd)$ are the transmittance through a sheet of thickness d for the case of direct and diffuse radiation, respectively. (This relationship is not exact because it does not account for the fact that the reflectance of a surface is dependent on angle. Furthermore, depending on the geometry of a spectrometer, integration in a second direction might be required.)

3.5.2.1 The Stokes' Formulas for an Assembly of Sheets

The equations that govern the passing of light through a sample composed of sheets are cumbersome, but they are readily solved with a computer. Let x represent the distance into the sample compared to the thickness of a single sheet. For example if x is 2, it is twice the thickness of a single sheet, and if x is $1/2$, it is half the thickness of a single sheet. From the fractions of a single layer (given by A_1 , R_1 , and T_1), we can calculate the fractions for a distance x (with $A_x + R_x + T_x = 1$) by

$$T_x = \frac{\Omega - \Omega^{-1}}{\Omega \Psi^x - (\Omega \Psi^x)^{-1}} \quad (3.76)$$

$$R_x = \frac{\Psi^x - (\Psi^x)^{-1}}{\Omega \Psi^x - (\Omega \Psi^x)^{-1}} \quad (3.77)$$

Ω and Ψ are defined by

$$\Omega = \frac{1 + R_1 + T_1^2 + \Delta}{2R_1} \quad (3.78)$$

and

$$\Psi = \frac{1 + R_1 + T_1^2 + \Delta}{2T_1} \quad (3.79)$$

where

$$\Delta = [(1 + R_1 + T_1)(1 + R_1 - T_1)(1 - R_1 + T_1)(1 - R_1 - T_1)]^{1/2} \quad (3.80)$$

Note that x can be any number, including a fraction. A real sample made up of sheets must of course have an integral number of sheets, so R_x for a fractional x does not correspond to anything physically observable, but it can be computed.

3.5.2.2 The Dahm Equation

When Dahm and Dahm [62,63] applied the discontinuous approach to the same problem as Kubelka and Munk, they made the following assumptions:

1. The sample has two large flat (but not necessarily smooth) plane parallel surfaces.
2. All radiation is directed, moving either forward or backward in a direction defined by the normal to the sample surface.

3. A sample is divided into layers. Each layer is representative of the sample as a whole and is nowhere more than one particle thick. The layer may contain voids, and there is no condition that the layer is the same thickness everywhere.
4. There is no scatter from one particle to another within the same layer.
5. The process in which radiation leaves a layer and returns into the hemisphere bounded by the plane of the sample's surface is termed remission. All remitted radiation is included in the description, whether its origin is regular reflection or backscatter.
6. Each layer will absorb a certain fraction, a , remit a certain fraction, r , and transmit a certain fraction, t .

Dahm and Dahm [62,63] have developed a discontinuum theoretical treatment based on the original work of Benford [67]. It is assumed that each layer of material is bounded by two parallel infinite planes. Of the radiation entering the i th layer, the total forward flux, that is, the fraction leaving the layer in the same direction, is t_i , and the total backward flux is r_i , and the fraction absorbed is a_i . The total forward flux includes both transmission and forward scatter and the total backward flux includes both external and internal reflection and backscatter.

Benford's equations allow the total forward and backward flux and the total absorbance to be calculated in terms of the properties of an individual layer. The overall properties for a material composed of two layers are given by

$$t_{i+j} = \frac{t_i t_j}{(1 - r_i r_j)}; \quad r_{i+1} = r_i + \frac{t_i^2 r_j}{(1 - r_i r_j)}; \quad a_{i+j} = 1 - t_{i+j} - r_{i+j} \quad (3.81)$$

If t_i , r_i , and a_i represent the known properties of a sample that contains a number i of identical layers, each of which is described by t_1 , r_1 , and a_1 , the properties of a sample containing i such layers is given by Benford as

$$t_{i+1} = \frac{t_i t_1}{(1 - r_i r_1)}; \quad r_{i+1} = r_i + \frac{t_i^2 r_1}{(1 - r_i r_1)}; \quad a_{i+1} = 1 - t_{i+1} - r_{i+1} \quad (3.82)$$

Dahm and Dahm considered what happens when the thickness of the sample is doubled or halved, and computed the total backward flux, R_∞ for an infinitely thick sample by an iterative solution of the latter set of equations. They showed that if the sample consists of an infinite number of layers, each with a forward flux of t_i , a backward flux of r_i , and absorption of a_i ,

$$f(R_\infty) = \frac{a_i}{2r_i} (2 - a_i - 2r_i) \quad (3.83)$$

The K–M equation was derived for a matrix of infinitesimally small particles. Because the numerator of the K–M function is the absorption coefficient, $f(R_\infty)$ varies linearly with the concentration of each component of the sample. However, it is known that the K–M equation is only an approximation. The Dahm Equation (3.83) may be expected to give a more exact solution for diffuse reflectance in the case of particles of finite size.

The Dahm equation is frequently expressed in terms of an Absorption–Remission function $A(R, T)$ which has as one of its characteristics that it has the same value for any thickness of a sample.

$$A(R, T) = \frac{(1 - R^2) - T^2}{R} = \frac{a}{r} (2 - a - 2r) \quad (3.84)$$

3.5.3 THE REPRESENTATIVE LAYER THEORY

3.5.3.1 Model for a Layer Representative of Particulate Solids

Use of the mathematics for plane parallel described in the last section requires that each layer in a sample has a single set of spectroscopic properties: an absorption fraction, a remission fraction, and a transmission fraction. This mathematics may be straightforwardly applied to homogeneous layers such as plastic sheets. In order to apply the mathematics to samples of particulate solids meaningfully, we need to establish a method for determining the properties of a layer of the sample from the properties of the individual particles.

In the model of the representative layer that we present here, each particle type is characterized by its composition, volume, and the average cross-sectional surface area that it presents to the incident beam. We will picture a particle as having two flat ends, each perpendicular to the incident beam. While the shape of the cross section is not important for our model, it may be pictured as a square.

Thus, a representative layer is made up of voids and of particles (shaped as boxes) of varying size. The ends of the boxes are perpendicular to the incident beam. The representative layer is of varying thickness, but is never more than a single particle thick. The layer is representative if

1. The volume fraction of each particle type is the same in the layer as in the sample as a whole.
2. For all particle types, the cross-sectional surface area in the layer is in the same proportion as the surface area of the particle type in the sample as a whole.
3. The fraction of the cross-sectional surface area that is made up of voids is the same as the void fraction of the sample as a whole.

With this model, it will be possible to calculate the properties of a representative layer from the properties of the individual particles of which it is comprised. The properties of the single particle that are of interest are the absorption, remission, and transmission of the incident beam.

3.5.3.2 Absorption and Remission of the Representative Layer

The absorption is a property of a molecule and can be well represented as a continuum, but remission is a property of an interface and may not be well represented as a continuum. In the absence of absorption, the remission is independent of the thickness of a layer. The remission fraction from a single representative layer is dependent only on the fraction of the cross-sectional surface area occupied by each type of particle and the remission power of the material of which the particle is composed. In the presence of absorption, the remission fraction diminishes (causing a reduction in the remission coefficient).

For cases where we know the thickness of the representative layer, it is possible to calculate the absorption and remission fractions for the representative layer from the remission and transmission fractions of any sample of known finite thickness, d . The absorption fraction, A_1 , of a layer is given by $[1 - \exp(-Kd)]$. (This is the value given by the plane parallel mathematics for a layer that has no remission. In the symbolism being used, the subscript refers to the number of layers. Thus A_2 would refer to the properties of two layers, not to second layer.) By implication, the absorption of a single particle, a , is given by $[1 - \exp(-Kd)]$.

A *remission coefficient* may be defined as the remission fraction of the representative layer divided by the thickness of the representative layer. For cases where the thickness of the representative layer is not known, the plane parallel mathematics can be used to obtain absorption and remission coefficients as described above. This requires an assumption that a sample can be well represented as a continuum. In this *linear* region, for a given linear absorption coefficient the absorption fraction of a particle is proportional to its thickness, and the following conditions will be observed (and will be seen in the

examples below to explain some commonly observed phenomena):

1. The contribution of a particle type to absorption is proportional to the volume fraction (including voids) of the particle type and to the absorption coefficient of the material making up the particle.
2. The contribution of a particle type to remission is proportional to the total cross-sectional surface area of the particle type in the representative layer and the remission power of the material making up the particle.
3. In a mixture of two or more particle types of similar remitting power, the Absorption–Remission function of each particle type is represented in the Absorption–Remission function of the sample weighted in proportion to surface area to volume ratio of the particle type.

In regions of higher absorption, the coefficients obtained from the plane-parallel mathematics become less reliable. Then the representative layer model is a better descriptor, though the assumptions about the shape of the particles become more important. The regions of extremely high absorption tend not to be of value for compositional analysis. Here, the contribution of a particle to the absorption of a layer is proportional to the surface area fraction of the particle type and not dependent on its thickness. The representative layer model would still be useful in applications such as image analysis.

3.5.3.3 Mathematical Expression of Model

A sample is made up of particles of various types. For each particle type, we have the following definitions of symbols:

d_i	the thickness of a particle of type i in the direction of the incident beam
ρ_i	density of a particle of type i
w_i	weight fraction of a particle of type i
v_0	void fraction of the sample
v_i	fraction of <i>occupied</i> volume in the sample composed of particles of type i
V_i	fraction of <i>total</i> volume occupied by particles of type i
s_i	fraction of particle surface area which belongs to particle type i
S_i	fraction of a cross-sectional surface comprised of particles of type i
k_i	the effective absorption coefficient of the particles of type i
b_i	the effective remission coefficient of the particles of type i
$(bd)_i$	the remitting power of the material comprising particle type i

Volume surface area and fractions, for a given particle type can be computed from weight fractions and particle density as follows:

$$v_i = \frac{w_i / \rho_i d_i}{\sum (w_i / \rho_i d_i)} \quad (3.85)$$

$$s_i = \frac{w_i / \rho_i}{\sum (w_i / \rho_i)} \quad (3.86)$$

The following formulas assume that the amount of transmitted light lost by an interaction with a single particle either to absorption or remission is small. With this assumption, for a single particle,

the fraction of light absorbed is given by the cross-sectional area and the Bouguer–Lambert law, and the remission fraction is given by the cross-sectional area and $b_i d_i$. Thus, for a representative layer

$$A = \sum S_i [1 - \exp(-k_i d_i)] \quad (3.87)$$

$$R = \sum S_i b_i d_i \quad (3.88)$$

When $k_i d_i$ is small for all particle types, the following approximation can be made:

$$A = \sum S_i k_i d_i$$

Then, since the surface area times the thickness is the volume,

$$A = \sum V_i k_i \quad (3.89)$$

This is the basis for statement 1 in the previous section.

It has been shown [3,68] experimentally that b_i is proportional to $1/d_i$. This is equivalent to saying that remission is proportional to surface area, and follows from the assumption that remission is a property of an interface. This implies that the product $b_i d_i$ is constant and a property of the composition of the material making up the particle, independent of particle size. The term $(bd)_i$ is defined as the *remitting power* of a material, and

$$R_i = \sum S_i (bd)_i \quad (3.90)$$

Note that Equation (3.90) is a special case of Equation (3.88), and is the basis for statement 2 in the previous section.

In the linear range, the approximation is reasonable that

$$A(r, t)_i = \frac{V_i a_i}{S_i r_i} \quad (3.91)$$

When all particles types are in this range, it follows that

$$A(R, T) = \frac{\sum a_i}{\sum r_i} \quad (3.92)$$

For the case where the remission fraction is the same for all particle types, $A(R, T)$ is equal to $\Sigma(V_j a_j / S_j r_j)$ which is the basis for statement 3 above.

3.6 APPLICATION OF THEORY TO MODEL SYSTEMS*

In this section, we will give several hypothetical examples, though we will compare results from the theoretical model to actual experimental data. One set of data is from a system investigated by Olinger and Griffiths [50]. Up to three components (carbazole, NaCl, and graphite) were mixed

* Section 3.6 is taken with minor modification by permission of NIR Publications Copyright 207 from: Donald J. Dahm and Kevin D. Dahm, *Interpreting Diffuse Reflectance and Transmittance: A Theoretical Introduction to Absorption Spectroscopy of Scattering Materials* pp 133–152, NIR Publications (2007). Sections 3.6.1 and 3.6.2 resulted from a collaboration between the authors, which was based on work described in Reference 50.

during the preparation of each sample. Carbazole was intended to represent a typical organic analyte having moderate absorption. It will be referred to as the analyte. NaCl and graphite respectively represent nonabsorbing and absorbing compounds. Whether as single compounds or mixtures, they will sometimes be referred to as the matrix. In the hypothetical examples, we will use the names of these materials, even though our calculation will not necessarily use the exact properties of these materials.

If we consider a beam of light encountering a single layer of the particles, there are several possible fates for the light:

1. The light can go through a void in the layer. In these examples, we will neglect the effect of voids, which tends to be minimal in a sample of close packed particles.
2. Light can interact with the surface of any of the particles and be reflected.
3. Light can enter the NaCl particles and a portion will be scattered backward. The balance of the light will be transmitted through the particle. In modeling this system, we will assume that there is no absorption by the NaCl particles.
4. Light can enter the graphite particles and all of it will be absorbed. The absorption of graphite is very high at most wavelengths in visible and infrared regions, so this assumption well approximates an actual property.
5. Light can enter a carbazole particle and a portion of the light will be absorbed. Another portion of it will be remitted, with the balance transmitted.

In total, the layer will remit a certain fraction, which represents the sum of reflection and backscatter; absorb a certain fraction; and transmit (either directly or diffusely) the balance.

One approach to modeling the system would be to use the absorption and remission (or scattering) coefficients with the following assumptions:

- The absorption coefficient of the NaCl is zero.
- The absorption coefficient of the graphite is infinite.
- The absorption coefficient of carbazole is neither zero nor infinite.

The linear absorption coefficient of carbazole might be determined by making transmission measurements on a solution of known concentration in a nonabsorbing solvent. Such a coefficient is a material property of carbazole. Unfortunately, that is not the absorption coefficient that is referred to in the theories of diffuse reflection. The theories refer to an absorption coefficient that would be obtained by taking the absorption and remission properties of a finite sample and extrapolating them to a sample of zero thickness. For a sample of zero scatter, the two absorption coefficients would be the same. In the presence of scatter, they are different. However, at very high absorption levels, the corresponding fraction of light absorbed as defined by each coefficient becomes indistinguishable experimentally. (This should not be interpreted as meaning that the two absorption coefficients become mathematically equal at very high absorption levels. The two coefficients will both approach infinity as absorption levels becomes very large, but will always have different values at the same absorption fraction, and will change at different rates as the absorption fraction changes.)

It is the intermediate levels of absorption that are the most important for analyses. Here, the relationship between the two coefficients is rather complex. No one chose this complicated situation: nature and our mathematics thrust it on us. However, many workers either did not know about or ignored the fact that the two coefficients are not proportional. Thus we urge care when reading the literature on this issue.

The situation with the remission coefficient is equally undesirable. Two particles with the same surface reflectivity and particle size will have different remission (or scattering) coefficients if the absorption coefficient of the material within the particle is different. These differences can be

illustrated by our extremes. If there is infinite absorption by the particle, the remission from the particle will be only the reflection from the first particle surface. For the particle with no absorption (but the same surface reflectivity), there will also be remission from the back surface of the particle. When extrapolated to zero thickness, this will mean that the remission coefficient for the particle with infinite absorption will be approximately half that of the case of zero absorption. (This is an over simplification. The remission coefficient will go down with increasing absorption and then, as the absorption becomes very large, will increase [69].) Just as for the case of absorption, many workers have ignored or did not know about the change in remission coefficients. (In the older literature, what we term a remission coefficient is called a scattering coefficient. The term scattering coefficient also has other meanings. Again, care is urged when reading the literature.)

There are situations where the assumptions of proportionality of the absorption coefficients and constancy of remission coefficient are reasonable. If the fraction of light absorbed by a single particle is small, the assumptions are good. In order to describe other situations, it is desirable to use an approach that does not suffer from the complexities of coefficients referred to above. In the following examples, we will illustrate the use of the approach of discontinuum theory as embodied in the representative layer theory.

Discontinuum theories are sometimes called particle theories. We will treat a sample as an assembly of particles. For the general case, we do not know how to use our models to deduce particle properties from the spectroscopic data. We can, however, use the models to calculate the spectroscopic properties of a sample from those of the particle.

3.6.1 EXAMPLE 1: GRAPHITE IN NaCl

We will begin our series of examples by assuming that we are mixing particles that have a drastically different absorption fraction but the same remission fraction. The plot in Figure 3.6a shows how $f(R_\infty)$ (denoted there as K-M) and a/r vary as a function of the particle fraction of graphite in a mixture of graphite (infinite absorption) and NaCl (zero absorption), with each assumed to have a remission fraction of 0.04. The value of 0.04 was chosen because it is the specular reflectance at normal incidence from the surface of a planar sample with a refractive index of 1.5. The value of $f(R_\infty)$ was calculated from the Dahm Equation (3.84). Because the function $A(R, T)$ is constant for

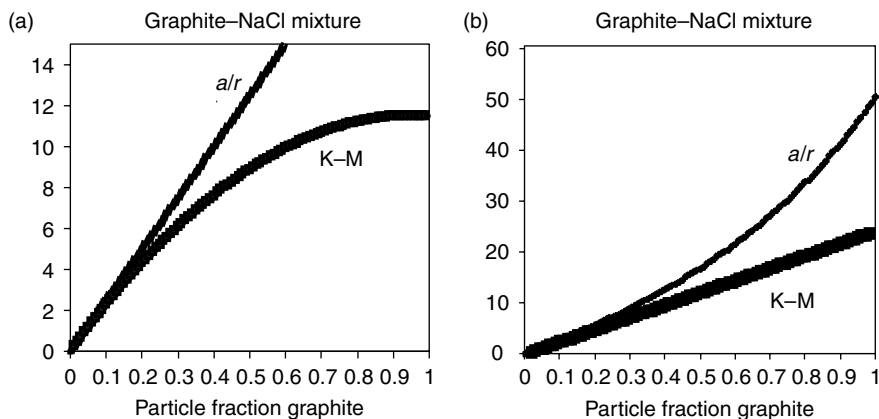


FIGURE 3.6 Variation of $f(R_\infty)$ (denoted here as K-M) and a/r as a function of the particle fraction of graphite in a mixture of graphite (infinite absorption) and NaCl (zero absorption). In chart (a), each is assumed to have a remission fraction of 0.04. In (b), the remission from the graphite is assumed to be half that from the NaCl. (Reproduced from D. J. Dahm and K. D. Dahm, *Interpreting Diffuse Reflectance and Transmittance: A Theoretical Introduction to Adsorption Spectroscopy of Scattering Materials* (2007), by permission of NIR Publications Copyright 2007.)

any thickness of the sample, we can calculate the value for infinite thickness ($T_\infty = 0$) from the value for the single layer.

$$f(R_\infty) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{A(R_\infty, 0)}{2} = \frac{a}{2r}(2 - a - 2r) \quad (3.93)$$

Here the upper case letters A , R , and T are used to denote the fractions of incident light absorbed, remitted, and transmitted by a sample of any thickness. The lower case letters are used to denote the corresponding fractions for a single layer or particle. For samples of infinite thickness, the K–M function, $f(R_\infty)$, and the Absorption–Remission function, $A(R, T)$, are related by a simple factor of two. The K–M equation, as originally derived, contains a quantity related to scatter in its denominator, which with the assumption of isotropic scatter, is twice the magnitude of r .

The value for the absorption fraction of a single layer, a , varies from 0 to 1 with the fraction of the area of the sample occupied by graphite, while the value for the remission fraction, r , is constant at 0.04 (i.e., 4% of the incident radiation is reflected from either NaCl or graphite; any radiation entering a graphite particle is completely absorbed; radiation entering a particle of NaCl may be scattered within the sample in either the forward or reverse directions). Notice that the ratio a/r increases linearly with graphite concentration in this example. We may picture this situation as the sequential replacement of a clear particle (with no absorption) with an opaque particle (with infinitely high absorption). The increase in a (the absorption fraction of the layer) is proportional to the number of particles we have replaced. However, the remission from the layer is constant by our assumption. On the extreme right of the chart, the layer is completely opaque and therefore all the remission comes from reflection from the front surface of the first layer. At the left of the chart, where the particles are clear, the remission is coming from many layers. The observation of the nonlinearity in the function $f(R_\infty)$ (also true for $\log(1/R_\infty)$) is qualitatively explained as being because the effective pathlength of the light in a sample decreases with increasing absorption.

Of course, we do not literally count out particles when we make up a sample; rather we weigh them. Because graphite and NaCl have approximately the same density (2.2 g/cm³), the particle fraction is the same as the weight fraction (assuming all particles have the same size). In later examples, we will encounter a situation where this is not true. Here we make the point that because they have the same density, particles of graphite and NaCl of the same size can be mixed to form a matrix having any desired absorption fraction. We will refer to such a matrix in Example 2.

Just above, we considered particles with vastly different absorptivities, and the same remission fraction. If we have two particles with the same refractive index but different absorptivity, the remission fraction will not be the same. The remission fraction from a highly absorbing particle (from which all the remission comes from the front surface) will be roughly half that of a lightly absorbing particle (where the remission comes from both the front and rear surfaces). It is the remission from the front surface that sets the upper limit on the K–M function.

Next, we will consider a similar system, but add that complexity. The graphite is assumed to have an absorption fraction of one and a remission fraction of 0.02, while the NaCl retains a remission fraction of 0.04 and zero absorption. The plot in Figure 3.6b shows how $f(R_\infty)$ (denoted there as K–M) and a/r vary as a function of the particle fraction of graphite in a mixture of graphite (infinite absorption, reflectance of 2%) and NaCl (zero absorption, reflectance of 4%). Notice that not only has the value of a/r changed, but it is also nonlinear, because $1/r$ is not a linear function. However, notice that the K–M plot in Figure 3.6b is linear. This is a surprising result given our contention expressed earlier that “reflection from the front surface of the sample leads to nonlinear plots of $f(R_\infty)$ against the concentration of an absorbing component in a nonabsorbing matrix, with $f(R_\infty)$ approaching its maximum value asymptotically.” If we consider the Dahm equation expressed as $f(R_\infty) = (a/2r)(2 - a - 2r)$, this behavior is explained by the nonlinearity in the ratio a/r observed in Figure 3.6b being exactly compensated by the decrease in r in the expression: $(2 - a - 2r)$. The assumptions that we made are reasonable for a system in which one particle absorbs

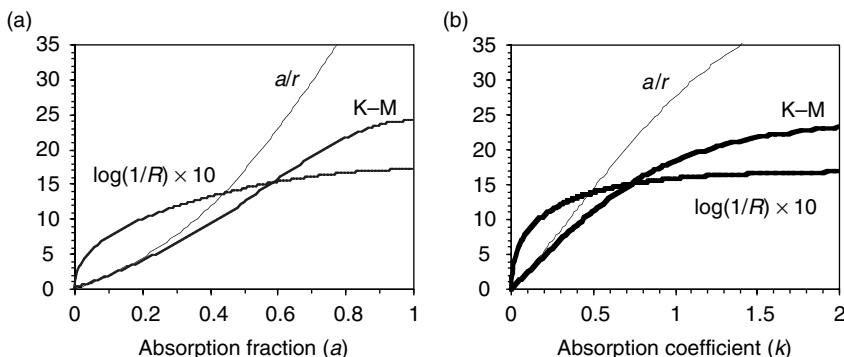


FIGURE 3.7 Functions related to a hypothetical sheet modeling particles within which the absorption fraction (a) is changing. The reflectivity (r) of the sheet surface is taken as 2%. All points plotted in chart (b) have a corresponding point in (a) with the same ordinate value, the difference between the two being the abscissa. The absorption coefficient (k) is taken as $[-\ln(1-a)]$. (Reproduced from D. J. Dahm and K. D. Dahm, *Interpreting Diffuse Reflectance and Transmittance: A Theoretical Introduction to Adsorption Spectroscopy of Scattering Materials* (2007), by permission of NIR Publications Copyright 2007.)

the light totally and the other not at all. However, this is not the normal situation encountered in NIR analyses.

In Figure 3.7, we consider a more normal case. We assume that all particles in a sample are identical. As the linear absorption coefficient k increases, a layer in each sample will have a progressively larger absorption fraction, which in the absence of scatter would increase linearly from 0 to 1. The collection of samples considered will have the same pattern in absorption fraction as those considered just above. The difference is that the increase is happening within a particle rather than by mixing particles with different absorption fractions. The remission fraction is calculated using the model of a Stokes sheet (described by Equation (3.67) to Equation (3.69)). The sheet surface has a reflectance of 2%, the same as that of the absorbing particles above. The remission fraction still decreases with absorption, but the decrease follows a more realistic pattern than in the earlier example. The decrease in remission again yields a nonlinear plot for a/r .

Notice that $f(R_\infty)$, labeled K-M in Figure 3.7, now varies nonlinearly with a and has a shape more typical of real samples. For the purpose of comparison, the plot of $\log(1/R_\infty)$ vs. a is also shown. The plots in Figure 3.6b and Figure 3.7 invite the conclusion that it is not only the surface reflectance in itself that causes the nonlinearity of the K-M function. It results from a rather complex interaction between the remission and absorption in a scattering sample.

The data in Figure 3.7b are the same data as in Figure 3.7a, but plotted to better mimic how the metrics vary as a function of composition. The abscissa in Figure 3.7b is the absorption coefficient (k) that would yield the absorption fraction (a) in Figure 3.7a using the formula $[k = -\ln(1-a)]$.

3.6.2 EXAMPLE 2: CARBAZOLE IN A MATRIX OF VARYING ABSORPTION

At the time the experimental work referred to at the beginning of this series of examples [50] was undertaken, there was a belief that *matrix referencing* should remove the effect of the matrix and isolate the absorption of the analyte. The experimental data (reported in previous editions of this Handbook) showed unambiguously that, whether using $f(R_\infty)$ or $\log(1/R_\infty)$ as the metric, the matrix referencing technique would not work in the way envisioned. However, when the experimental work was done, there was not a theory available that would predict the results that were obtained.

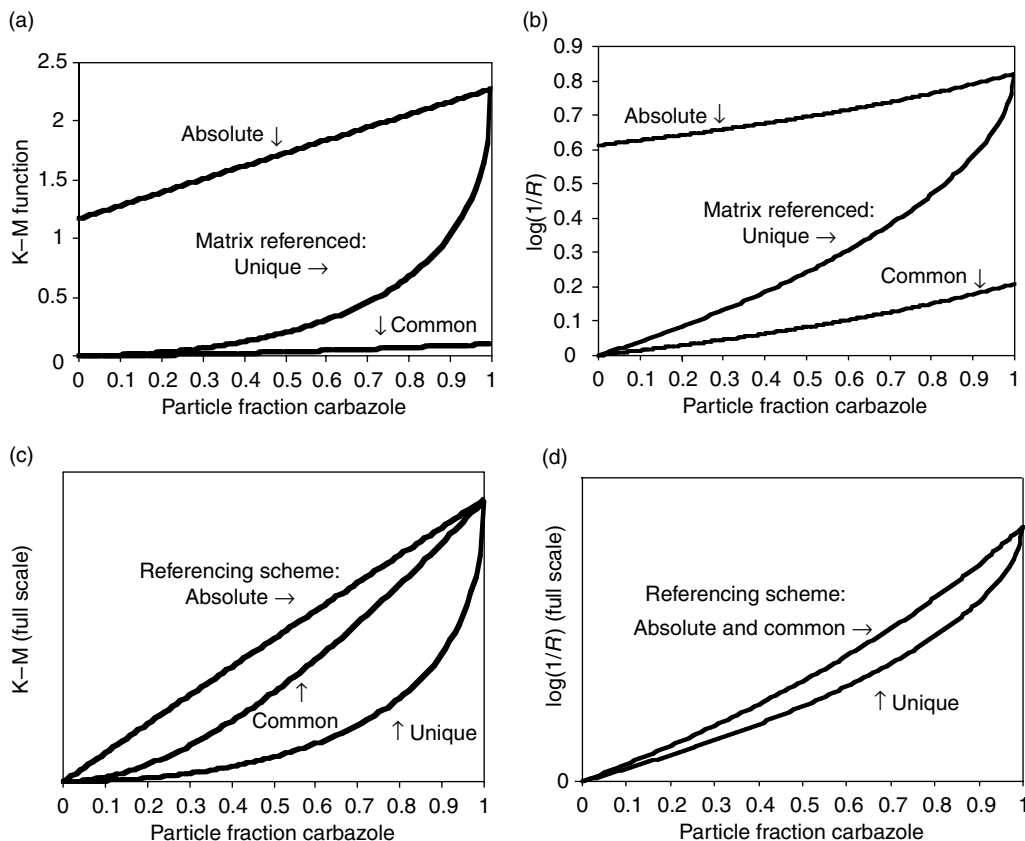


FIGURE 3.8 Variation of $f(R_\infty)$ and $\log_{10}(1/R)$ with the particle fraction of carbazole for a matrix composed of 95% NaCl and 5% carbon for three cases: (a) a referencing scheme that puts the data on an absolute scale (marked “Absolute” on figure); (b) a reference that is made up of the matrix, each particle of which has an absorption fraction of 0.05 (5%) of the light incident upon it (marked “Common” in figure); and a (c) matrix referencing scheme in which a unique reference is made for each sample (marked “Unique” in figure). Over relatively small ranges of analyte absorption, matrix absorption does not cause significant nonlinearity of K–M data collected on an absolute scale. Matrix referencing introduces nonlinearity into the data. In general, the $\log(1/R'_\infty)$ data is not linear, as can be seen in chart (d). (Reproduced from D. J. Dahm, *NIR news*, 15: 6–10 (2007), by permission of NIR Publications Copyright 2007.)

In this example, we will investigate the effects of various referencing techniques with the intent of explaining the experimental results. The four panels in Figure 3.8 each exhibit the behavior for a matrix composed of 95% NaCl and 5% carbon for all three cases. The three cases are for

1. A referencing scheme that puts the data on an absolute scale (marked “Absolute” in Figure 3.8).
2. A reference that is made up of the matrix, each particle of which has an absorption fraction of 0.05 (5%) of the light incident upon it (marked “Common” in Figure 3.8).
3. A matrix referencing scheme in which a unique reference is made for each sample (marked “Unique” in Figure 3.8). These reference materials are assumed to have the same contribution of matrix absorption in the reference (uniquely made for this sample) as in the sample. Thus if the absorption fraction for a matrix layer is 0.05 (5%), and the particle fraction of the matrix is 0.5 (50%), then the contribution of the reference material to the absorption fraction of a single layer of the sample would be $(0.05 \times 0.5 = 0.025)$ or 2.5%.

This was the referencing technique that was explored in the referenced experimental work [50].

In each case, we have assumed that the absorption fraction of the carbazole is 0.10 (10%), and that of the matrix is 0.05 (5%). The remission fraction from all particles is assumed to be 0.04 (4%).

Calculations according to the representative layer theory show that the matrix referencing techniques would not make a plot of $f(R_\infty)$ or $\log(1/R_\infty)$ more linear than data on an absolute scale. In fact, matrix referencing has the opposite effect. Notice that in Figure 3.8a, the K–M function for data on an absolute scale, for this case, varies approximately linearly with the particle fraction of carbazole. In Example 1, the corresponding function was very nonlinear. This is because here we assumed a maximum change of 5% in the absorption of a particle instead of 100%, which was the case assumed for graphite in Example 1. (This will make it clear that it is the referencing scheme that is causing the nonlinearity.) A plot of $\log(1/R_\infty)$ vs. the particle fraction of carbazole in Figure 3.8b is slightly less linear than the corresponding plot for the K–M function.

In an article by Griffiths on continuum methods of diffuse reflection in the previous edition of this *Handbook*, it was pointed out that “converting the measured reflectance of carbazole to the K–M functions yields a fairly linear plot when the carbazole analyte was dispersed in the nonabsorbing matrix (NaCl).” Consistent with this observation, the representative layer theory predicts that in this case, the unique matrix referencing would give the same plot as the line marked “Absolute” in Figure 3.8c. This is because the absorption for NaCl is very low, and thus in an infinitely thick sample, is nearly 100% reflective. Conversely, when the matrix is an absorbing one, the experimenters concluded “that $\log(1/R'_\infty)$ values provided a more linear plot over a major portion of the concentration range studied.” That conclusion seems very reasonable in light of a comparison between the lines marked “Unique” in Figure 3.8c and Figure 3.8d.

To subtract the matrix absorption successfully, two criteria must be fulfilled: (a) the reference must match the absorption of the matrix and (b) the remission from the reference must remain constant at all concentrations. In this matrix-referencing scheme, we have assured that we have matched the absorption of the matrix and kept the remission *for a single layer* constant at all concentrations. The remission of the reference will vary with concentration because the effective pathlength changes with absorption.

Figure 3.8c and Figure 3.8d show the same data as Figure 3.8a and Figure 3.8b, except that all plots are displayed full scale to show the relative linearity of each case more clearly. The K–M data which has all been referenced to a material with the same absorption (calculating the function $f(R'_\infty)$, where $[R' = R_\infty/R_{\text{ref}}]$) has a different shape than the $f(R_\infty)$ data that is on an absolute scale. The log data has the same shape, and the curves fall upon each other when expanded to full scale. This behavior is not unexpected, because taking the logarithm of a ratio is equivalent to subtracting the logs. The correct way to correct K–M data for a reference with nonzero absorption is to subtract the K–M functions, not to take their ratio. The software packages accompanying many NIR instruments do not do this. This tends to make $\log(1/R)$ a better choice for an absorption metric than K–M on those instruments for artificial reasons.

From the preceding example, it can be seen that the representative layer theory predicts that the unique matrix referencing technique would not work. In this example, we will show that the shape of the experimental curves is predicted by the representative layer theory. In this study, $\log(1/R)$ and K–M matrix referenced data were obtained from three series of infinitely thick samples:

1. Binary mixtures of carbazole and NaCl.
2. Carbazole dispersed in a matrix of NaCl and 1 wt.% graphite.
3. Carbazole dispersed in a matrix of NaCl and 5 wt.% graphite.

Using the techniques described in the preceding examples, the curves shown in Figure 3.9a and Figure 3.9b were obtained for $f(R_\infty)$ and $\log(1/R)$, respectively. Notice the only curve for which

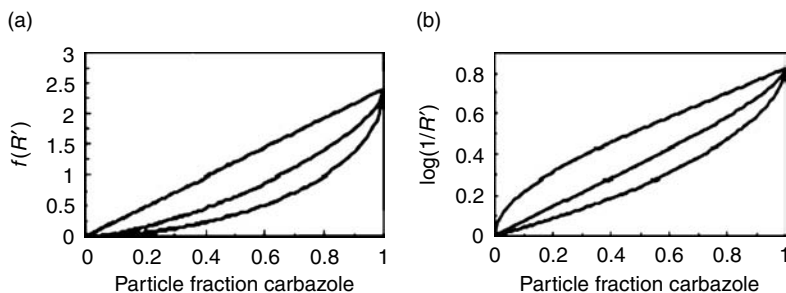


FIGURE 3.9 A unique matrix referencing scheme can yield a linear K–M plot only with a nonabsorbing matrix and when plotted against a volume based measure of concentration. The $\log(1/R_\infty)$ curves are in general not linear. From top to bottom, the lines in each chart represent 0, 1, and 5% absorption by matrix particles. The lines represent values calculated by the representative layer theory. (Reproduced from D. J. Dahm, *NIR News*, 15: 6–10 (2007), by permission of NIR Publications Copyright 2007.)

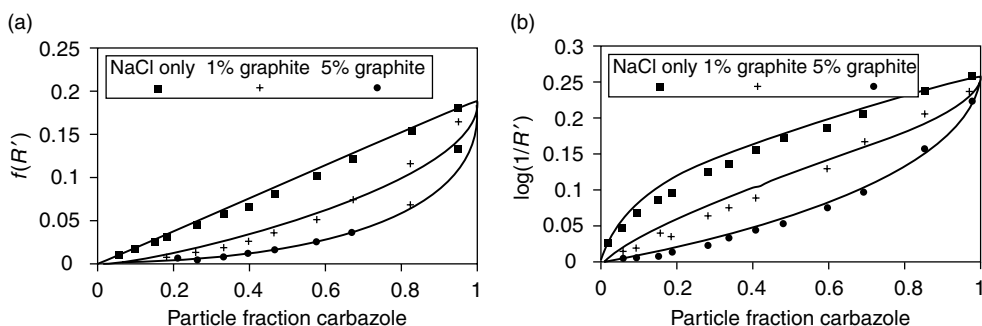


FIGURE 3.10 A unique matrix referencing scheme can yield a linear K–M plot only with a nonabsorbing matrix and when plotted against a volume based measure of concentration. From top to bottom, the lines in each chart represent 0, 1, and 5% absorption by matrix particles. The markers are experimental data. The lines are fit to the points assuming that the K–M function would be linear for all plots if an absolute reference were used. (Reproduced from D. J. Dahm, *NIR News*, 15: 6–10 (2007), by permission of NIR Publications Copyright 2007.)

a straight line is predicted is when $f(R_\infty)$ is measured under conditions where the matrix does not absorb. This line is straight because the absorption of the matrix is assumed to be zero, and thus the data for that series are on an absolute scale. If all data were on an absolute scale, the data in all three curves would form straight lines ending at the 100% carbazole absorption, although each would begin at a different point on the y-axis.

The data calculated from the representative layer theory may be compared with actual data shown in Figure 3.10. These data were originally reported by Olinger and Griffiths [50] and modified by changing the weight fraction to particle fraction, as described by Dahm and Dahm [70]. The individual points are experimental data, while the lines in these plots were calculated assuming that $f(R_\infty)$ varied linearly with the volume concentration of both the analyte and matrix. Notice the similarity between the shapes of the experimental curves and those in Figure 3.9a and Figure 3.9b, which are calculated from the representative layer theory. The absolute values of the numbers are different from the experimental data for a variety of reasons, one being that the experimental data was base line adjusted. More important, as we shall see in the next example, is that the particle properties assigned which give this fit are not necessarily those of real particles.

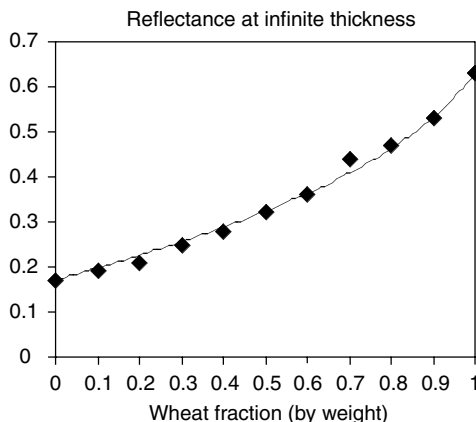


FIGURE 3.11 Reflectance measured (diamonds) by image analysis using visible light compared to reflectance calculated (line) from the Dahm equation for mixtures of wheat and rape seed meal. (Reproduced from D. J. Dahm and K. D. Dahm, *Near-Infrared Technology in the Agriculture and Food Industries*, 2nd edn. (P. Williams and K. Norris, eds.), St. Paul, MN, pp. 1–17 (2001) by permission of American Association of Cereal Chemists, Inc. Copyright 2001).

3.6.3 EXAMPLE 3: MIXTURE OF WHEAT AND RAPE SEED MEAL

Additional experimental confirmation for the usefulness of the representative layer theory comes from the work of DeVaux et al. [71,72]. This study was concerned with the fact that, in a mixture of two components with different particle sizes, the smaller particles will be *over-represented* in an absorption metric compared to the weight fractions of the components. This effect was systematically examined with a combination of image analysis and NIR spectroscopy on mixtures of “fine and coarse fractions of wheat and rape seed meal.”

In order to be able to characterize the surfaces of the mixtures by image analysis with visible radiation, mixtures were made from “raw materials of contrasting colors. A white product and a black product were chosen: wheat and rape seed meal.” Four series totaling forty samples were built from “mixtures from 0 to 100% of wheat by steps of 10%: fine wheat with fine rape seed meal, fine wheat with coarse rape seed meal, coarse wheat with fine rape seed meal, and coarse wheat with coarse rape seed meal.” Figure 3.11 shows experimental data (as diamonds) for the mixtures of fine wheat with fine rape seed meal (i.e., passed through the same size sieve). The remission of visible radiation from the mixtures increased with the concentration of “white” wheat. The reflectance values were “not placed on a straight line between the raw fractions.”

To explain the shape of an image analysis curve, it is natural to think in terms of an opaque layer, some parts of which are less reflective (darker) than others. In this example, the dark areas are assumed to reflect 0.1 (10%) of the light and absorb 0.9 (90%); while the bright areas reflect 0.9 (90%) and absorb 0.1 (10%). The reflectance is shown for this hypothetical opaque layer in Figure 3.12a. Notice that the reflectance, R , is a straight line proportional to the fraction of bright area (as it would be for any mixture composed of two components of different reflectance); as a result, the K–M and absorbance functions are both curved.

However, the reflectance from a real layer of white wheat and dark rape seed is not opaque. Light is transmitted through the white particles, interacts with interior layers of the sample, and reemerges from the sample after having undergone additional absorption. In Figure 3.12b, we see a plot of the image analysis data from mixtures of the dark rape and white wheat seed meal. Here the line representing the reflectance is curved, as is the line representing the K–M function. The $\log(1/R)$ curve, by coincidence, happens to be approximately straight in this case.

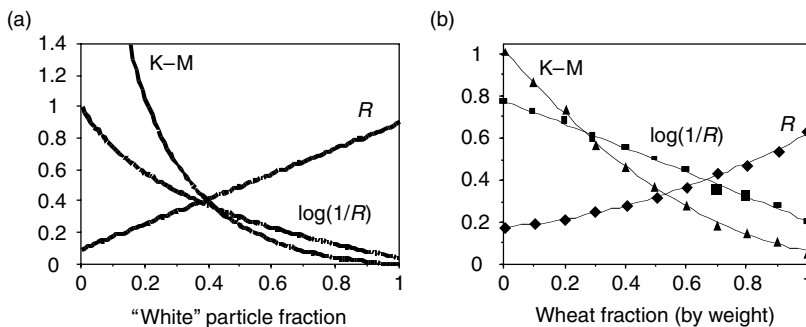


FIGURE 3.12 Variation of R , $f(R_\infty)$ and $\log_{10}(1/R)$ with particle fraction of the lighter component. (a) Calculated values when the dark areas are assumed to reflect 0.1 (10%) of the incident radiation and absorb 0.9 (90%) and the bright areas reflect 0.9 (90%) and absorb 0.1 (10%). Notice that the reflectance, R , is a straight line proportional to the fraction of bright area (as it would be for any mixture composed of two components of different reflectance); as a result, the K–M and absorbance functions are both curved. (b) The reflectance from a real layer of white wheat and dark rape seed. Light is transmitted through the white (wheat) particles, interacts with interior layers of the sample, and reemerges from the sample after having undergone additional absorption. Here the line representing the reflectance is curved, as is the line representing the K–M function. The $\log(1/R)$ curve, by coincidence, happens to be approximately straight in this case. (Reproduced from D. J. Dahm and K. D. Dahm, *Interpreting Diffuse Reflectance and Transmittance: A Theoretical Introduction to Adsorption Spectroscopy of Scattering Materials* (2007), by permission of NIR Publications Copyright 2007.)

Focusing on the line where the absorption metric is reflectance (the fraction of the incident beam that is remitted), the black rape seed particles have high absorption, while the wheat particles absorb far less. However, the line is curved because the layer is not opaque as discussed previously. Further, because the effective sample thickness is higher for wheat than for rape seed, the line is super-linear as a function of wheat fraction. As in Example 2, the representative layer theory can reasonably define the shape of these curves.

In the model used to fit these experimental data, we assumed that the remission at the left-hand side of the chart is from an opaque layer, although the opaque layer of pure rape seed is almost certainly more than one particle thick because of the effect of voids. In a sample with voids, an opaque layer must be at least two particles thick. The fractions for a hypothetical layer were calculated from the assumed characteristics of the individual components using Equation (3.84) and Equation (3.85). The R_∞ values for an infinitely thick sample making up the line in Figure 3.12 were calculated by applying an inverse form of the Dahm equation

$$R_\infty = \frac{1 + R_1^2 - T_1^2 - ((1 + R_1^2 - T_1^2)^2 - 4R_1^2)^{0.5}}{2R_1} \quad (3.94)$$

The hypothetical opaque layer that fits yields the desired R_∞ value for 100% rape seed; 0% wheat has the following parameters: $a_r = 0.83$, $r_r = 0.17$, $t_r = 0.00$. With these values for the pure rape seed layer, the following values for a pure wheat layer: $a_w = 0.089$, $r_w = 0.43$, $t_w = 0.48$ (subscript w for wheat, r for rape seed) gave the desired line.

These assigned values may be quite unreasonable for remission from a layer that contained nothing but those particles. This is a curve fitting exercise. We are manipulating four parameters to fit a line. There are many combinations of these four parameters that give the same line. Presumably, none of the sets will be the properties of a single layer of particles, because we have not included voids in the model, and a layer of particles, one particle thick, would certainly contain voids. Nonetheless, this exercise illustrates the capability of the Dahm equation and the representative layer theory to describe particulate systems, even when there is very high absorption, a place where there

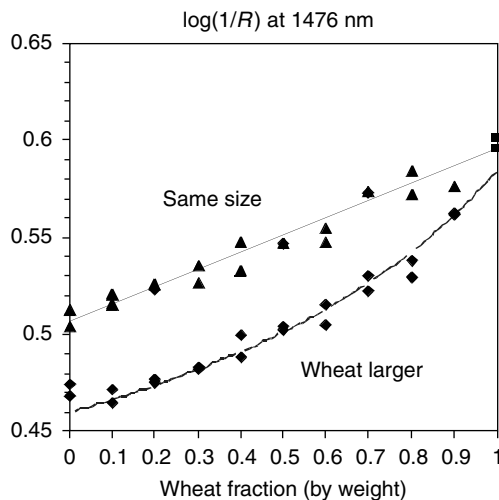


FIGURE 3.13 Variation of $\log(1/R_\infty)$ vs. the fraction of wheat in a sample composed of rape seed and wheat. The markers show the experimental values of $\log(1/R_\infty)$ vs. the fraction of wheat in the mixture while the lines show the shape predicted by the representative layer theory. The top curve is for wheat and rape seed of the same particle size. In the bottom curve, the wheat particles are twice the size of the particles of rape seed. (Reproduced from D. J. Dahm and K. D. Dahm, *Near-Infrared Technology in the Agriculture and Food Industries*, 2nd edn. (P. Williams and K. Norris, eds.), St. Paul, MN, pp. 1–17 (2001) by permission of American Association of Cereal Chemists, Inc. Copyright 2001.)

is significant departure from the K–M equation. However, this is not a method to determine straightforwardly the properties of an individual particle from the experimental data. We can not show the quantitative validity of the theory unless we know the properties of the individual particles (from other data).

In Figure 3.13, we show the effect of particle size on the spectroscopic data. We have not engaged in any curve fitting exercise. The data is taken from a region of the NIR spectrum in which absorption of both rape seed and wheat is low. This will assure that absorption levels will not cause a significant deviation from linearity of a plot of $\log(1/R)$ vs. wheat fraction. The wavelength of 1476 nm was chosen because it is the location of an absorption maximum for rape seed, even though the absorption at this wavelength range is low, as indicated by $\log(1/R)$ values of less than 0.6.

For the $\log(1/R)$ data at 1476 nm measured for two series of mixtures shown in Figure 3.13, the upper series has wheat and rape seed with the same particle size. Notice that, as expected at low absorption levels, the experimental points generally follow a straight line. The lower series has wheat particles that are twice as large as the rape seed. Here the points show a departure from linearity. The end points of the line were determined by human intervention, while the shape of the line was determined from the values for the end points weighted in proportion to the contribution of each component to the total surface area in the sample. (Surface area is inversely proportional to particle size, so the smaller particles are over-represented compared to their weight fraction.) The two series have the points at the extreme right of the chart in common. Again the data are reasonably well fit by calculations using the representative layer theory. However, this time we have not used the Dahm equation or done any curve fitting.

This data set also illustrates the problems in making reproducible measurements on mixtures of particles. The largest source of variation is probably the void fraction. The effect of void fraction on absorption is profound in close-packed mixtures, but the changes tend to occur at all wavelengths and are thus easily corrected.

Notice how the end point of the series with two particle sizes departs from the trend for the central points. Samples with only one particle size do not pack as densely as one with two sizes. In a

mixture with two sizes, the small particles ones fill in the holes between the large ones. The behavior of such samples may be understood with use of the Dahm equation (Equation (3.84)). As the void fraction decreases, both a and r increase and the factor $(2 - a - 2r)$ decreases and the value of the Absorbance/Remission function, $A(R, T)$, decreases along with it.

We believe these examples show both the usefulness of the representative layer theory in describing particulate samples. Further we have illustrated the applicability of the Dahm equation to situations where continuous theories do not work well. Finally, we have tried to make clear that, however useful the theory may be in explaining what we observe, we have not created an easy way to determine particle properties, or to correct simply absorption data so it will be linear with analyte concentration.

3.7 EXPERIMENTAL CONSIDERATIONS FOR REFLECTION MEASUREMENTS

Most analyses performed by NIR spectroscopy rely heavily on a chemometric approach. Most of these rely on some form of linear regression. The recommendations made here are guided by the assumption that the more linear the metric of reflectance as a function of analyte concentration presented to the chemometric package, the better the performance. When one is choosing an experimental arrangement, convenience generally outweighs the quest for optimal linearity. Additionally, there is no known approach that will produce linear data for the general case. However, there are some principles that we would like to extend:

1. For small particles (where “small” means that the product of linear absorption coefficient and diameter is much less than one), plots of $f(R_\infty)$ vs. c are linear for samples of infinite thickness. Achieving this linearity requires either that absorption by the matrix is very low or measurements are made using absolute reflectance.
2. For nonscattering samples placed on a specularly or diffusely reflecting substrate, $\log(1/R)$ is linear for measurements made in transfection, since this is simply a double-pass transmission measurement. For scattering samples measured in transfection, plots of $\log(1/R)$ vs. the concentration of a given component, are more linear for thinner samples than thicker ones. The most linear response is obtained when measurements are made in regions of low to moderate absorption.

The optimal experimental arrangement, of course, will depend on the purpose of the measurement, as well as the characteristics of the sample of interest. The discussions below are intended to help the analyst make better decisions in selecting an experimental arrangement.

3.7.1 DEPTH OF PENETRATION

If diffuse reflection spectrometry is to be used for quality control purposes, it is essential to know the actual investigated sample volume, which is equivalent to the radiation penetration depth or the effective sample size m_{eff} . In a publication by Berntsson et al. [73], the effective sample size of pharmaceutical powders was investigated by the three-flux approximation presented above that they called the equation of radiative transfer (ERT) method. and an empirical method they called the variable layer thickness (VLT) method. In this publication, the effective sample size m_{eff} is defined as the mass per area of the sample at which its diffuse reflectance has reached 98% of the diffuse reflectance of a corresponding optically thick sample.

To obtain m_{eff} with the use of the three-flux approximation of ERT, the scattering and absorption coefficients of the investigated powder are obtained from diffuse reflectance and transmittance measurements on optically thin samples (for a detailed discussion see [41,74]). Then, at each wavelength, the diffuse reflectance is calculated as a function of the scattering and absorption coefficients for a

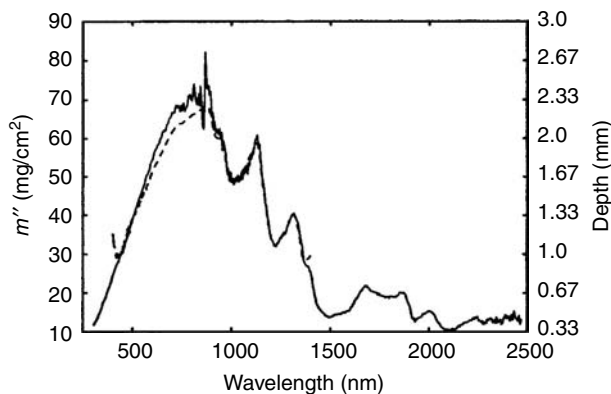


FIGURE 3.14 Effective sample size and penetration depth of microcrystalline cellulose powder according to the ERT method (solid line) and the VLT method (dashed line). (Reproduced from O. Berntsson, T. Burger, S. Folestad, L.-G. Danielsson, J. Kuhn, and J. Fricke, *Anal. Chem.*, 71: 617–623 (1999) by permission of American Chemical Society Copyright 1999.)

gradually increasing sample thickness and compared to the diffuse reflectance of an infinitely thick sample. The sample thickness for which the above 98% limit is reached corresponds to the effective sample size m_{eff} .

This method is compared to a totally independent procedure, the VLT method, where diffuse reflection spectra are collected at several controlled powder thicknesses. Assuming that the backing is either completely transparent or completely opaque, the diffuse reflectance of the powder layers increases with increasing sample thickness until the reflectance of an optically thick sample (R_{∞}) is reached. For each measured wavelength, an exponential function is fitted to the experimental data (plots of $\log(1/R)$ vs. sample thickness, where R is the measured reflectance). Using the 98% limit, the effective sample size m_{eff} can be obtained from the exponential fit (see [75] for a detailed discussion of the VLT method).

Figure 3.14 presents the effective sample size of a microcrystalline cellulose powder (MCC, particle size range: 65 to 300 μm) in the NIR region. Assuming a powder density of 0.30 g cm^{-3} , the effective mass per area is transformed into an effective penetration depth in millimeters, shown on the right axis. The wavelength dependence of the effective sample size, which is inversely correlated to the absorption coefficient, and the good correspondence of the two methods can clearly be seen. The upper limit for the VLT method was 1400 nm for the MCC sample because the curve fitting becomes unstable if the effective sample mass per area is below the smallest powder mass per area used in the measurements. However, both methods are suited to determine the actual probed sample volume of a diffuse reflection measurement.

Another approach to the determination of penetration depth in terms of “number of particles interrogated” was investigated by Olinger and Griffiths [50]. An estimate of the effective penetration depth of the beam in a diffuse reflectance measurement can be made from the measured values of $\log(1/R'_{\infty})$. If the absorptivity of a pure material in the absence of scatter is given by k , and particles of the material have an average diameter d , then the absorbance per particle is given by $a'd$, where a' is the base 10 absorptivity. By dividing the baseline corrected values of $\log(1/R'_{\infty})$ by the absorbance per particle, the effective number of particles through which the radiation reaching the detector has passed can be estimated. Olinger and Griffiths estimated the typical number of particles through which the remitted beam passes at 20. More recent work by a number of different research groups suggests that the number of particles through which the remitted beam passes may be larger than 20, but that the sample depth that is actually interrogated in a diffuse reflection measurement is no larger than about 20 particles.

3.7.2 EFFECT OF RELATIVE REFLECTANCE AND MATRIX REFERENCING

Even when we are dealing with “small” particles for which there is no significant inherent deviation from linearity of plots of $f(R_\infty)$, experimental conditions can cause a nonlinear response. Further, certain approaches that are used by experimenters, borrowed largely from transmission spectroscopy, designed to improve the linearity, actually create a nonlinear response. Such is the case with “matrix referencing.” Below we wish to make two major points:

1. Using a relative reflectance measurement will tend cause a departure from linearity when using the K–M function, but not $\log(1/R)$, as a metric for absorption.
2. Matrix referencing will cause a nonlinearity for both metrics (if the matrix has absorption).

An important point in practical NIR measurements is that a relative reflectance, R'_∞ , rather than an absolute diffuse reflectance, R_∞ , is usually measured. The relative reflectance is equal to the ratio I_S/I_R , where I_S is the intensity of radiation reflected from the sample and I_R is the intensity of radiation reflected from a reference material. (In the case of NIR reflectance spectrometry this material is usually a ceramic disk while a powdered alkali halide is usually employed for mid-infrared measurements.) Strictly speaking, the K–M equation as well as the other presented models require that an absolute reflectance, defined as the ratio I_S/I_0 , be measured, where I_S is defined as above and I_0 is the intensity of the incident radiation. Absolute intensities can, in theory, be measured by using an infinitely thick sample of a scattering, but nonabsorbing, material as the reference. (The ideal material would have the same scattering properties as the sample.) The measurement of absolute reflectance can be made directly through the use of an integrating sphere, although some care is required. An absolute reflectance can also be derived from the relative reflectance although the derivation requires several experimental measurements [2].

The determination of absolute reflectance is more important when using the K–M function as the metric for absorption than for $\log(1/R)$ data. If we define (I_S/I_R) as the relative reflectance R' , then $\log(1/R') = \log(I_R/I_S) = \log(I_R) - \log(I_S)$. The reference value $\log(I_{\text{ref}})$ is a constant offset at a particular wavelength. Furthermore, if the reference is the matrix in which the analyte is embedded, the measurement becomes a direct measure of the absorption of the analyte. However, the K–M function using relative reflectance is not linearly related to the K–M function using absolute reflectance, as shown in the examples in Section 3.6. Matrix referencing with the K–M function would require measuring the function for the matrix and the sample, each on an absolute scale and subtracting.

Kortüm [2,76] showed experimentally that the use of relative reflectance causes deviations from K–M theory if the analyte is surrounded by an absorbing matrix. This effect was shown by measuring the visible reflectance of Cr_2O_3 in an absorbing matrix relative both to the pure diluent and to a highly reflective standard (MgO). When the diluent was used as the reference, a plot of $f(R'_\infty)$ vs. concentration of Cr_2O_3 was nonlinear. When MgO was used as the reference, a similar plot yielded a straight line with a positive intercept. Kortüm states that a straight line with a zero intercept can be obtained when $f(R_{\infty,A})$ is plotted vs. c :

$$f(R_{\infty,A}) = \frac{K_A}{S} = f(R'_{\infty,A+M} \times \rho R'_{\infty,M}) - f(\rho R'_{\infty,M}) \quad (3.95)$$

where $R_{\infty,A}$ is the absolute reflectance of the analyte (A), $R'_{\infty,A+M}$ is the reflectance of the analyte + matrix relative to the reflectance of the matrix (M), ρ is the absolute reflectance of a nonabsorbing standard, and $R'_{\infty,M}$ is the reflectance of the matrix relative to the reflectance of the nonabsorbing standard. For most NIR reflectance analyses of cereal products, it is impossible to measure either the value $R'_{\infty,M}$ or $R'_{\infty,A+M}$ due to the complexity of the sample. Therefore this type of correction for the

absorption of the matrix should only be viable for samples that are relatively simple, for example, binary and ternary mixtures. Even for simple samples, use of Equation (3.95) may not be productive since treatment of this type usually does not linearize plots of $f(R'_{\infty})$ vs. c for powdered mixtures measured by mid-infrared diffuse reflectance spectrometry [49], because in that case the absorption is usually very strong, and the function is inherently nonlinear even when absolute intensities are used.

It is generally accepted that the K–M equation (like Beer’s law, but to an even greater extent) is a limiting equation and should only apply for weakly absorbing bands, that is, when the product of absorptivity and concentration is low. For organic materials, absorptions in the NIR are due to vibrational overtones and combination bands. The absorptivities of these bands are much weaker than the absorptivities of the corresponding fundamental vibrations. Thus most organic analytes can be considered to be weakly absorbing in the NIR even without dilution. As noted above, however, for most NIR analyses, the analyte (such as protein or lipid molecules in a cereal) is usually not isolated from other components, but is surrounded by a matrix which is not only complex but which also often absorbs the incident radiation at least as strongly as the analyte at the analytical wavelengths. In a cereal product analysis the matrix would largely consist of carbohydrate molecules. It would therefore be expected that unless a proper referencing method is used, absorption by the matrix surrounding the analyte will cause deviations from the K–M equation. Even if the intensities are properly measured, using matrix referencing (for the case of an absorbing matrix) can cause a deviation from linearity.

3.8 CONCLUSIONS

In summary, therefore, although detailed investigations of the theory of diffuse reflection spectrometry by many workers have been carried out during the last century, they have not resulted in a single metric that is proportional to analyte concentration (in the same manner as absorbance in transmission spectroscopy). Fortunately, simple conversion of reflectance values to $\log(1/R'_{\infty})$ appears to be effective for many powdered samples being analyzed by NIR diffuse reflection spectrometry.

Because of the limitations inherent in representing a real sample as a continuum, the more sophisticated radiative transfer models, such as the discrete ordinate approximation or the diffusion approximation, hold little hope for obtaining for a better understanding of the effects occurring in diffuse reflection spectrometry for the general case.

Beginning with the properties of the individual particles in a mixture, the representative layer theory gives a way of calculating the properties of a layer of particles. The merging of the continuous and discontinuous approaches, as embodied in the Dahm equation, gives a way to calculate the spectroscopic properties of a sample from that of such a layer.

The inverse problem, calculating the properties of individual particles from that of a sample remains unsolved for the general case.

ACKNOWLEDGMENT

One of us (PRG) is grateful for financial support from the Department of the Army, Edgewood Chemical Biological Center under Contract Number DABJ05-03-P-1272.

REFERENCES

1. J. H. Lambert, *Photometria sive de mensura et gradibus luminis colorum et umbrae*, Augustae Vindelicorum (1760).
2. G. Kortüm, *Reflectance Spectroscopy*, Springer-Verlag, New York (1969).

3. W. W. Wendlandt and H. G. Hecht, *Reflectance Spectroscopy* (P. J. Elving and I. M. Kolthoff, eds.), Interscience, New York, (1966).
4. H. Wright, *Ann. Physik*, 1: 17 (1900).
5. G. I. Pokrowski, *Z. Physik*, 30: 66 (1924).
6. G. I. Pokrowski, *Z. Physik*, 35: 35 (1926).
7. G. I. Pokrowski, *Z. Physik*, 36: 472 (1926).
8. G. P. Woronkoff and G. J. Pokrowski, *Z. Physik*, 20: 358 (1924).
9. G. Mie, *Ann. Physik*, 25: 377 (1908).
10. H. C. van de Hulst, *Light Scattering by Small Particles*, John Wiley and Sons, New York (1957).
11. G. Oster, *Chem. Rev.*, 43: 319 (1947).
12. M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation*, (Ernest M. Loebl, ed.), Academic Press, New York, pp. 27–31 (1969).
13. J. W. Rayleigh, *Philos. Mag.*, 12: 81 (1881).
14. J. W. Rayleigh, *Philos. Mag.*, 47: 375 (1899).
15. H. C. van de Hulst, *Light Scattering by Small Particles*, Dover Publications, New York (1981).
16. H. H. Theissing, *J. Opt. Soc. Am.*, 40: 232 (1950).
17. S. Chandrasekhar, *Radiative Transfer*, Clarendon Press, Oxford (1950) (reprinted by Dover Publications, 1960).
18. J. S. Truelove, *J. Quant. Spectrosc. Radiat. Transfer*, 39: 27–31 (1988).
19. E. P. Incropera, T. D. Craig and H. G. Houf, *J. Quant. Spectrosc. Radiat. Transfer*, 31: 139–147 (1984).
20. A. Schuster, *Astrophys. J.*, 21: 1 (1905).
21. P. Kubelka and F. Munk, *Z. Tech. Physik*, 12: 593 (1931).
22. P. Kubelka, *J. Opt. Soc. Am.*, 38: 448 (1948).
23. N. C. Chaffin and P. R. Griffiths, *Appl. Spectrosc.*, 52: 218–221 (1998).
24. N. T. Melamed, *J. Appl. Phys.*, 34: 560 (1963).
25. Z. Bodo, *Acta Phys. Hung.*, 1: 135 (1951).
26. J. Broser, *Ann. Physik*, 5: 401 (1950).
27. E. O. Hulbert, *J. Opt. Soc. Am.*, 33: 42 (1943).
28. H. E. I. Neugebauer, *Z. Tech. Physik*, 18: 137 (1937).
29. M. Gurevic, *Physik. Z.*, 31: 753 (1930).
30. D. B. Judd, *J. Res. Natl. Bur. Std.*, 12: 354 (1934).
31. D. B. Judd, *J. Res. Natl. Bur. Std.*, 13: 281 (1934).
32. G. W. Ingle, *ASTM Bull.*, 116: 32 (1942).
33. T. Smith, *Trans. Opt. Soc. (London)*, 33: 150 (1931).
34. L. Amy, *Rev. Optique*, 16: 81 (1937).
35. H. D. Bruce, *Tech. Pap. 306, Natl. Bur. Std.* (1926).
36. L. Silberstein, *Philos. Mag.*, 4: 1291 (1927).
37. J. W. Ryde, *Proc. R. Soc. (London)*, A90: 219 (1931).
38. J. W. Ryde, *J. Soc. Glass Technol.*, 16: 408 (1932).
39. S. Q. Duntley, *J. Opt. Soc. Am.*, 32: 61 (1942).
40. M. G. Kaganer, *Opt. Spectrosc.*, 26: 240–242 (1969).
41. T. Burger, J. Kuhn, R. Caps and J. Fricke, *Appl. Spectrosc.*, 51: 309–317 (1997).
42. T. Burger, H. J. Ploss, J. Kuhn, S. Ebel and J. Fricke, *Appl. Spectrosc.*, 51: 1323–1329 (1997).
43. J. Kuhn, S. Korder, M. C. Arduini-Schuster, R. Caps and J. Fricke, *Rev. Sci. Instrum.*, 64: 2523–2530 (1993).
44. P. S. Mudgett and L. W. Richards, *Appl. Opt.*, 10: 1485–1502 (1971).
45. A. Kienle, M. S. Patterson, N. Dögnitz, R. Bays, G. Wagnières and H. van den Bergh, *Appl. Opt.*, 37: 779–791 (1998).
46. H. Wabnitz and H. Rinneberg, *Appl. Opt.*, 36: 64–74 (1997).
47. M. S. Patterson, E. Schwartz and B. C. Wilson, *Proc. SPIE*, 1065: 115–122 (1989).
48. R. G. Giovanelli, *Opt. Acta*, 2: 153 (1955).
49. P. R. Griffiths and M. P. Fuller, in *Advances in Infrared and Raman Spectroscopy*, vol. 9, Chapter 2 (R. J. H. Clark and R. E. Hester, eds.), Heyden Publishing Co., London, pp. 63–129 (1982).
50. J. M. Olinger and P. R. Griffiths, *Anal. Chem.*, 60: 2427–2435 (1988).
51. P. J. Brimmer and P. R. Griffiths, *Anal. Chem.*, 58: 2179–2184 (1986).

52. E. Pitts, *Proc. Phys. Soc. (London)*, Sect. B, 67: 105 (1954).
53. H. G. Hecht, *Anal. Chem.*, 48: 1775–1779 (1976).
54. H. G. Hecht, *Appl. Spectrosc.*, 34: 157–160 (1980).
55. H. G. Hecht, *Appl. Spectrosc.*, 30: 610–614 (1976).
56. G. V. Rozenberg, *Uspe. Fiz. Nauk*, 69: 666 (1959).
57. E. S. Kuznetsov, *Isv. Akad. Nauk. SSSR Ser. Geogr. Geofiz.*, 5: 247 (1943).
58. H. G. Hecht, *Appl. Spectrosc.*, 34: 161 (1980).
59. A. A. Christy, J. E. Tvedt, T. V. Karstang and R. A. Velapoldi, *Rev. Sci. Instrum.*, 59: 423 (1988).
60. M. L. E. TeVrucht, *Appl. Spectrosc.*, 43: 1492 (1989).
61. P. R. Griffiths, *J. Near Infrared Spectrosc.*, 3: 60 (1996).
62. D. J. Dahm and K. D. Dahm, *Appl. Spectrosc.*, 53: 647 (1999).
63. D. J. Dahm and K. D. Dahm, *J. Near Infrared Spectrosc.*, 7: 47–53 (1999).
64. N. T. Melamed, *J. Appl. Phys.*, 34: 560–570 (1963).
65. E. L. Simmons, *J. Appl. Phys.*, 46: 344–349 (1975).
66. G. G. Stokes, *Proc. Roy. Soc. (London)*, 11: 545 (1860/1862).
67. F. Benford, *J. Opt. Soc. Am.*, 36: 524–554 (1946).
68. G. Kortüm, W. Braun and G. Herzog, *Angew. Chem. Intern. Ed.*, 2: 333 (1963).
69. D. J. Dahm, K. D. Dahm and K. H. Norris, *J. Near Infrared Spectrosc.*, 8: 171–181 (2000).
70. D. J. Dahm and K. D. Dahm, *J. Near Infrared Spectrosc.*, 3: 53–59 (1995).
71. M.-F. DeVaux, N. Nathier Dufour, P. Robert and D. Bertrand, *Appl. Spectrosc.*, 49: 84–91 (1995).
72. D. J. Dahm and K. D. Dahm, Chapter 1 in *Near-Infrared Technology in the Agriculture and Food Industries*, 2nd edn. (P. Williams and K. Norris, eds.), The American Association of Cereal Chemists, St. Paul, MN, pp. 1–17 (2001).
73. O. Berntsson, T. Burger, S. Folestad, L.-G. Danielsson, J. Kuhn and J. Fricke, *Anal. Chem.*, 71: 617–623 (1999).
74. T. Burger, P. Keym, J. Kuhn and J. Fricke, *J. Near Infrared Spectrosc.*, 6: 33 (1998).
75. O. Berntsson, L.-G. Danielsson and S. Folestad, *Anal. Chim. Acta*, 364: 243–251 (1998).
76. G. Kortüm, *Naturwissen*, 47: 600–609 (1966).
77. B. McKellar and M. A. Box, *J. Atmos. Sci.*, 38: 1063–1068 (1981).
78. L. G. Henyey and J. L. Greenstein, *Astrophys. J.*, 93: 70–83 (1941).

Part II

Instrumentation and Calibration

4 Commercial NIR Instrumentation

Jerome J. Workman and Donald A. Burns

CONTENTS

4.1 Introduction	67
4.2 Instrumentation Design	68
4.3 Commercial Instrumentation	75
References	78

4.1 INTRODUCTION

Prior to World War II, the near-infrared (NIR) region was not considered to be particularly useful for organic compositional spectroscopy. The explanation for this line of reasoning seemed obvious. The NIR region consisted only of overtone and combination bands of fundamental molecular vibrations. It was also obvious that all these combination and overtone bands occur in a relatively narrow region (750–3,000 nm) as compared to the fundamental bands occurring at 2,800–50,000 nm. Thus it was observed that NIR bands are severely overlapped, difficult to resolve, and, once resolved, difficult to interpret. If the same information is obtained with better resolution in the infrared (IR) region, then why would any chemist be interested in the difficulties inherent with NIR spectroscopy?

The difficulties in using the NIR region for qualitative much less quantitative analysis seemed enormous. If samples were not dried prior to NIR analysis, the changes in hydrogen bonding due to the effects of sample temperature, ionic strength, and analyte concentration would complicate the interpretation of over-lapping NIR spectral bands. Changes in hydrogen bonding bring about real and apparent band shifts as well as flattening or broadening of band shapes. The overtone and combination molecular absorptions found within the NIR region inherently have significantly lower intensity as compared to the fundamental IR absorptions. Thus the changes in absorbance in the NIR region would be quite small with respect to changes in concentration. The relatively small extinction coefficients due to combination and overtone NIR bands would place severe restrictions on the allowable noise levels and stability for any NIR instrument in order to be useful for quantitative work.

With the expansion of technology and manufacturing capabilities in the early 1940s, the use of spectrophotometry gained a greater respect from chemical analysts. Commercial instrument manufacturers responded by offering a wide range of research quality instruments that could scan make analog photometric measurements in the ultraviolet (UV) (190–350 nm), visible (350–750 nm), NIR (750–2500 nm), and classical mid- and far-IR regions. The late 1940s and early 1950s brought about the proliferation of research papers describing the use of IR instruments for synthetic chemical and natural product identification. For the reasons described previously, the NIR region was considered less suitable for this type of work than the mid- and far-IR regions.

In the late 1960s, work led by Karl Norris, considered by many to be the “father” of NIR, demonstrated the potential value of this region for quantitative work by making measurements of agricultural products. The U.S. Department of Agriculture, the then-employer of Mr. Norris, made

a decision to provide research facilities for continued work in rapid measurement of agricultural products by using NIR. The only commercially available NIR instruments at that time were those developed to optimize measurements in the UV and visible regions, with the NIR capability only as an added feature. These instruments were noisy by modern standards and would not make adequate measurements without modifications. Whetsel [1] had characterized the essential features of seven manufacturers, all claiming to have commercially available instrumentation suitable for high-quality NIR measurements. These seven instrument companies included Beckman, Cary (now Varian), Coleman, Perkin-Elmer, Shimadzu, Unicam, and Zeiss.

Norris decided that because none of the commercially available instruments was designed specifically for making NIR measurements, he and coworkers would design a system. The practical use of such an instrument involved the application of correlation algorithms to account for the interferences in natural product spectra. The broad absorption bands for ground wheat and soybean spectra were extremely complex. Following this basic work, the Illinois Department of Agriculture solicited bids for Norris-designed instruments to measure protein, oil, and moisture in soybeans.

The first commercial unit was produced by Dickey-John. This instrument incorporated a tungsten-halogen source, six research grade interference filters, and uncooled lead sulfide detectors at 0 to 45° geometry. The samples measured on such an instrument had dry matters above 85% and were ground to pass a 1 mm screen prior to packing in a quartz-windowed sample cup. The unit was demonstrated at the 1971 Illinois State Fair. Following the introduction of this first instrument, Neotec designed and built a rotating filter grain quality analyzer. Both the Dickey-John and Neotec instruments were dedicated systems with analog circuitry; neither instrument was generally considered user-friendly.

In the mid-1970s, Technicon Instruments entered the arena by asking Dickey-John to manufacture the InfraAlzer instrument line for them. The first Technicon/Dickey-John instrument was named the Infrazyler 2.5 and included the features of “dust-proof” sealed optics and internal temperature control. These features improved the ruggedness and long-term wavelength stability of the instrument. Technicon also utilized the integrating sphere to give both reference reading and signal integration for each measurements. Neotec countered with improvements in their own systems. The competition was initially very intense to win a large order for instrumentation from the Federal Grain Inspection Service (FGIS). The FGIS had adopted NIR initially as an approved method of analysis. The result of the competition was orders for all three companies. It was noted that the ultimate performance of the initial instruments was roughly equivalent. For a detailed discussion on the history of NIR instrumentation, refer to Reference 2.

4.2 INSTRUMENTATION DESIGN

Instruments can be categorized in several ways to specified in Table 4.1. Instrument manufacturers will point out the advantages of their own instrument brands, but the consumer must be as objective as possible when making a purchasing decision. All instrument engineering designs either are optimized for a specific type of sample presentation geometry, or the design becomes a compromise system expected to solve a multiplicity of measurement problems. In the research phase an instrument with broad capabilities is generally the instrument of choice. Once an application has been clearly defined, an instrument optimized for the specific sample presentation geometry and sample type would give the most reliable results. A list of the criteria most often used to compare instrumentation is shown in Table 4.2.

Figure 4.1 shows the two most prevalent basic instrument designs common in modern NIR analysis. The simple differences can be quite dramatic when making spectroscopic measurements. For example, reflectance measurements penetrate only 1 to 4 mm of the front surface of ground samples. This small penetration of energy into a sample brings about greater variation when measuring nonhomogeneous samples than transmittance techniques.

TABLE 4.1
Distinguishing Characteristics of NIR Instrumentation (Hardware Only)

- 1.0 Optical Configurations
 - A. Interference filters (wedge, discrete)
 - B. Moving diffraction grating(s) (holographic, stigmatic)
 - C. Prism(s)
 - D. Near-infrared emitting diodes (NIR-EDs)
 - E. Interferometers (Michelson, etc.)
 - F. Acoustooptical tunable filters (AOTF)
 - G. Fixed grating (Hadamard mask exit slit with fixed grating)
 - H. Diode array detector (fixed grating)
 - I. MEMS (micro-electro-mechanical systems) tunable optical filter (MEMS-NIR)
 - 2.0 Scan Rates
 - A. Slow (60–90 s, or longer, for full range)
 - B. Medium (10–60 s for full range)
 - C. Fast (0.1–10 s for full range)
 - D. Ultra-fast 0.01–0.1 s for full range (single scan)
 - 3.0 Source Type
 - A. Infrared (Globar, Nernst glower, Nichrome wire, tunable laser)
 - B. Near-infrared (tungsten–halogen monofilament, NIR-ED)
 - 4.0 Detector Type
 - A. Infrared thermal type (thermocouples, thermistors, pneumatic devices)
 - B. Infrared photon detectors (semiconductors such as InAs, InSb, PbS, and PbSe)
 - C. Near-infrared photon detectors (semiconductors such as InGaAs, PbS, and InAs)
 - 5.0 Sample Averaging Technique (to average sample presentation effects)
 - A. Spinning or rotating cup
 - B. Stepwise averaging of long-path cell
 - C. Integrating sphere
 - 6.0 Dustproofing (for use in environments with fine dust)
 - A. Sealed optical chamber
 - B. Outer cabinet with O-ring seal
 - C. Positive air pressure within instrument chamber
 - 7.0 Waterproofing (for use in humid environments)
 - A. Internal drying feature (such as dry air or desiccant)
 - B. Sealed optics and electronics
 - C. Dry air positive air flow
 - 8.0 Vibration-tolerant
 - A. Low-mass moving parts
 - B. No moving parts
 - 9.0 Optimized Instrument Design for:
 - A. Noncontact reflectance (telescoping optics)
 - B. Near-infrared transmittance (NIT)
 - C. Near-infrared reflectance (NIR)
 - D. Reflectance and/or transmittance (NIR/NIT)
-

In transmittance measurements the entire pathlength of a sample is integrated into the spectral measurement, thereby reducing errors due to nonhomogeneity of samples. Transmittance techniques are most useful for measuring large particles. For fine particles, the front surface scatter brings about a loss of energy transmitted through a sample with the net effect being a decrease in the signal-to-noise of the instrument. In transmittance, higher frequency energy is most commonly used due to its

TABLE 4.2**Basic Instrument Specifications Used for Instrument Comparisons**

Wavelength range	The total <i>useful</i> wavelength range for the specified instrument. The useful range is limited by the specific lamp source used, the optical design, and the detector(s)
Bandwidth	For grating instruments, this characteristic is formally defined as the convolution of the dispersion of the grating and the slit functions for both the entrance and exit slits. This term is not equivalent to resolution but determines the resolution of an instrument. Bandwidth is sometimes equated to slit width or bandpass. The bandwidth or bandpass can be defined as the full width at half maximum (FWHM) of the bandshape of monochromatic radiation passing through a monochromator. With high-quality optics, the smaller the bandwidth, the higher the resolution
Wavelength accuracy	Wavelength accuracy for any spectrometer is defined as the difference between the measured wavelength of a wavelength standard and the nominal wavelength reported for that wavelength standard. Standard reference materials (SRMs) are available from the National Institute for Standards and Technology (NIST) for wavelength accuracy measurements
Photometric accuracy	Photometric accuracy is the difference between the nominal absolute photometric value in transmittance, reflectance, or absorbance of a standard reference material and the actual photometric measurement obtained using a specified instrument under test. SRMs for photometric accuracy measurements are also available from NIST. Photometric accuracy measurements are made under prespecified sample presentation geometries
Photometric noise	This is the actual root mean square deviation for sample scans for a specified wavelength range and for a specified absorbance level
Photometric range	This is the photometric range of an instrument from the highest useful absorbance that can be reproduced to the lowest detectable absorbance. This specification is also termed <i>dynamic range</i>
Baseline flatness	Every spectrophotometer will exhibit a change in the flatness of its baseline measurement. The degree of flatness in transmittance, reflectance, or absorbance is equal to the difference between the minimum and maximum values for a baseline reference scan. Baseline flatness is more precisely defined as the maximum difference in photometric values of a regression line drawn through all the data points of a baseline reference scan. A different term, <i>baseline drift</i> , refers to the change in photometric value of a spectrometer's baselines at a specific wavelength with respect to time
Stray light	Stray light, sometimes termed <i>stray radiant energy</i> , is the major cause of nonlinearity for most instruments. It is defined as the sum total of any energy or light other than the wavelength of interest that reaches the sample and detector
Scan speed(s)	Slow, medium, and fast scan speeds are usually defined in the most advanced UV-Vis-NIR spectrophotometers. In systems dedicated for NIR work, the scan speed may not be adjustable, due to the requirement for extremely low noise for NIR quantitative measurements
Digital smoothing	The analog signal received from the detector(s) is often processed differently between instrument manufacturers. Some instruments may smooth the data prior to exporting; different algorithms may be used to perform this smoothing. Other manufacturers may export raw data only
Data interval	This is the interval that data can be reported in or captured. The smaller the interval, the greater the information available for processing, deconvolution, etc
Scan modes	Many instruments allow the user to switch from transmittance to reflectance
Optics	The optical component quality determines the overall quality of measurements made on an instrument. Poor-quality optics result in high stray light, small dynamic range, higher noise, poor photometric accuracy and reproducibility, poor wavelength accuracy and reproducibility, etc

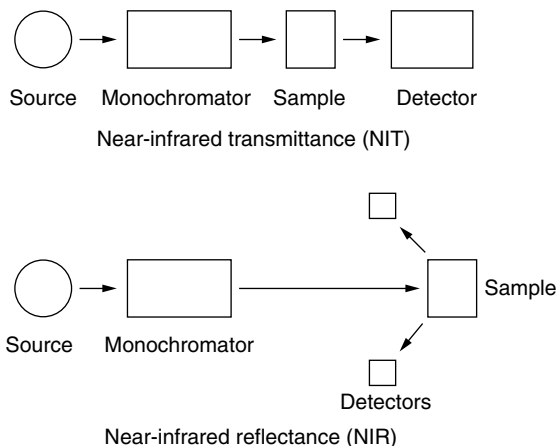


FIGURE 4.1 Basic instrument configurations.

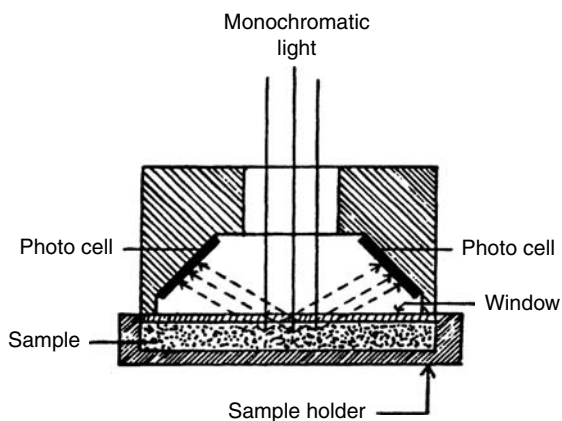


FIGURE 4.2 Diffuse reflectance using a 0 to 45° sample presentation geometry. (Reprinted with permission from K. Norris, *Reflectance Spectroscopy in Modern Methods of Food Analysis*, AVI, Westport, CT, 1984.)

greater depth of penetration into the sample. The higher frequency energy (800 to 1400 nm) is more susceptible to front surface scattering than lower frequency energy. Transmittance measurements must therefore be optimized, taking into consideration the relationships between the frequency used for measurement, front surface scatter, and the pathlength of the sample. In transmittance measurements, particle size can be small enough to begin to scatter most of the energy striking the sample. If the particle size is sufficiently small, the instrument will not transmit enough energy through the sample for the detectors to record a signal. To compensate, the ideal research instrument would have both transmittance and reflectance capabilities. Other chapters in this book discuss the advantages of the various presentation geometries with respect to specific applications.

Figure 4.2 is a diagram of an instrument detector system used for diffuse reflectance. This geometry provides for monochromatic light to illuminate the sample at a 0° angle (normal incidence) to the sample. The collection detectors for the reflected light are at 45°. Two or four detectors at a 45° angle can be used. Generally only lead sulphide (PbS) detectors are used for measurements in the 1100- to 2500-nm region, whereas PbS “sandwiched” with silicon photodiodes are most often used for visible-near-infrared work (typically 400 to 2600 nm).

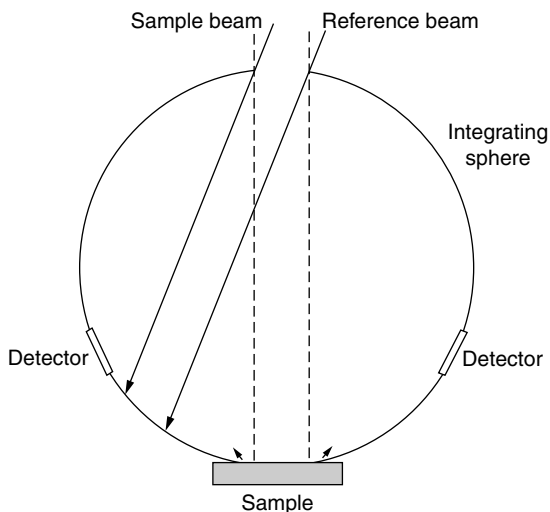


FIGURE 4.3 Diffuse reflectance using an integrating sphere sample presentation geometry.

The signal from the detectors is added into a low-noise, high-gain amplifier and then converted from analog to digital. The digital signal is exported from the instrument to an onboard or external microcomputer for data processing, calibration, and storage. The computer records a signal representing the actual wavelength used for measurement with the raw reflectance or transmittance digital data. This function is repeated for both the sample and the reference. The spectrum, then, is the difference between the raw reflectance measurement of the sample and the raw reflectance measurement of the reference material. Raw reflectance is converted to absorbance using the function $\text{Absorbance} = -\log(10) * \text{Reflectance}$, commonly referred to as $\log 1/R$. Raw transmittance is converted to absorbance using the expression $\log 1/T$.

Another sample presentation geometry quite common in NIR spectroscopic measurements is the integrating sphere (Figure 4.3). The use of integrating spheres dates back to the first commercial photometers. The greatest advantage of the integrating sphere in the early days is that a detector placed at an exit port of the sphere was not susceptible to energy fluctuations from the incident beam because of deflection (scattering), refraction, or diffraction of light when working in the transmittance mode. So-called “sweet spots” on photomultiplier tubes and early semiconductor and photodiode detectors made it impossible to get reproducible measurements using detectors without the benefit of an integrating sphere. As detector technology improves, the advantages of the integrating spheres for energy collection are not as critical. The use of a sphere provides for internal photometric referencing producing a sort of pseudo-double-beam instrument. Single-beam instruments must be set up to measure a reference material before or after the sample scans are taken. This requires some slight inconvenience on the part of the user. Whether the sampling geometry user spheres or 0 to 45°, there is no clear-cut advantage of one type of sample presentation system over the other. The main difference is that the 0 to 45° geometry lends itself better to a transmittance measurement than the integrating sphere systems (Figure 4.1).

The first Neotec filter-type NIR spectrophotometers utilized a tilting filter concept (Figure 4.4a). This concept utilized wedge interference filters that transmitted energy at varying wavelengths and bandpasses dependent on the incident angle of the light passing through the interference filter wedge. This concept was refined into the spinning filter systems (Figure 4.4b). The basic principle remained the same, but now the filters were mounted in an encoder wheel for greater positioning accuracy (wavelength reproducibility) and greater reliability.

Another type of filter system consisted of the interference filter instrument (Figure 4.5). This instrument type uses prespecified research grade discrete interference filters. The filters are mounted

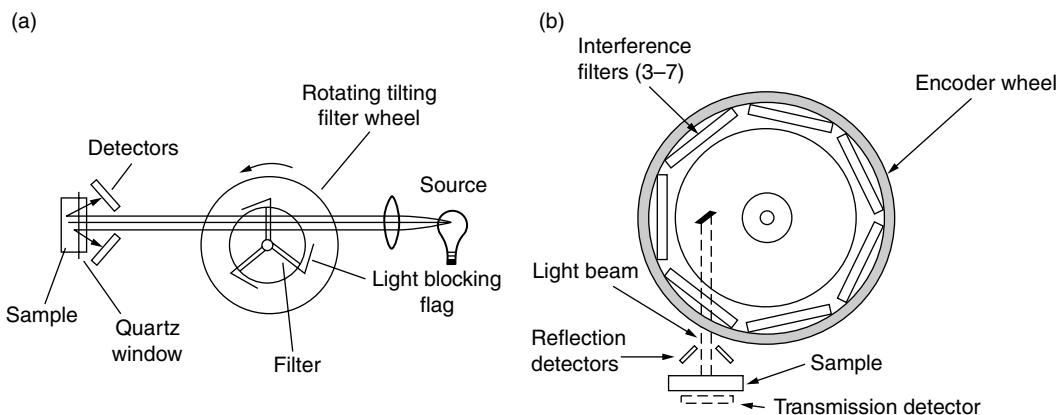


FIGURE 4.4 Two types of tilting filter instrument designs.

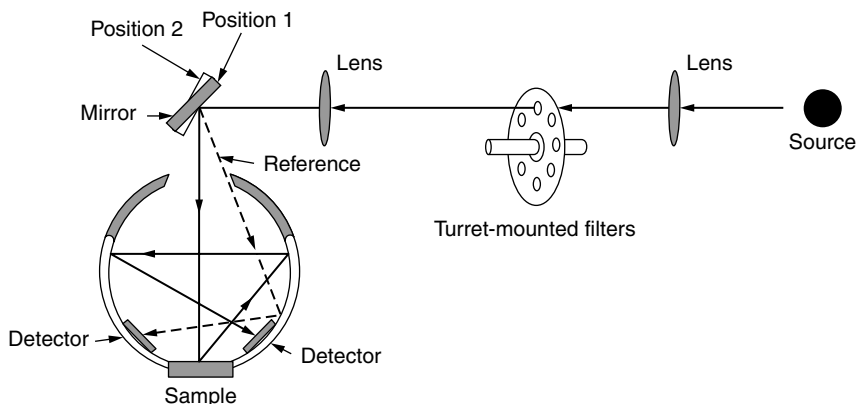


FIGURE 4.5 Typical turret-mounted discrete filter instrument. (Courtesy of Bran and Luebbe Analyzing Technologies, Inc.)

in a turret and rotated slowly to different positions during measurement scanning. These instruments are rugged but sometimes lack all of the information necessary for optimizing calibrations. If the selection of interference filters is not correct for the specific application where the instrument is required, successful calibrations are not possible. Systems currently exist configured with anywhere from 6 to 44 discrete wavelength interference filters.

Dedicated dispersive (grating-type) scanning NIR instruments have been available since 1978. These instruments varied in optical design, yet all systems had the common features of tungsten-halogen source lamps, single monochromator with a holographic diffraction grating, and uncooled PbS detectors. The design of a system that became available in 1982 is shown in Figure 4.6. This complex instrument was typical of the slower scanning systems and has a rather complex optical diagram for a single monochromator instrument. Note the integrating sphere that gives the enhanced feature of a pseudo-double-beam operation.

The modern designed monochromators are shown in Figure 4.7 and Figure 4.8. These instrument designs differ in that Figure 4.7 shows a typical predispersive monochromator-based instrument where the light is dispersed prior to striking the sample. In Figure 4.8, the postdispersive design allows the transmission of more energy into either a single fiberoptic strand or a fiberoptic bundle. The white light is piped through the fiberoptic where it strikes the sample, and returns to the dispersive

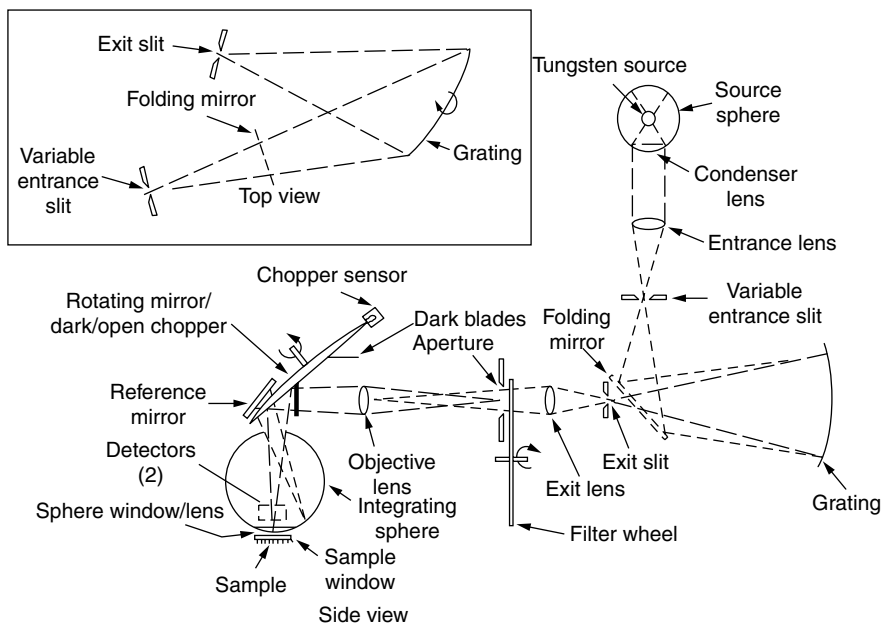


FIGURE 4.6 Optical design of a single monochromator scanning instrument. (Courtesy of Bran and Luebbe Analyzing Technologies, Inc.)

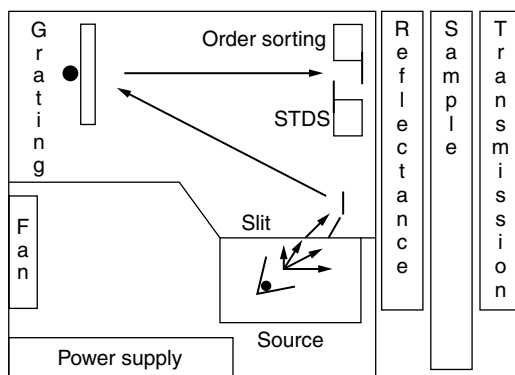


FIGURE 4.7 Diagram of a predispersive, fast-scanning, grating spectrophotometer configured for reflectance or transmittance. (Courtesy of NIR Systems, Inc.)

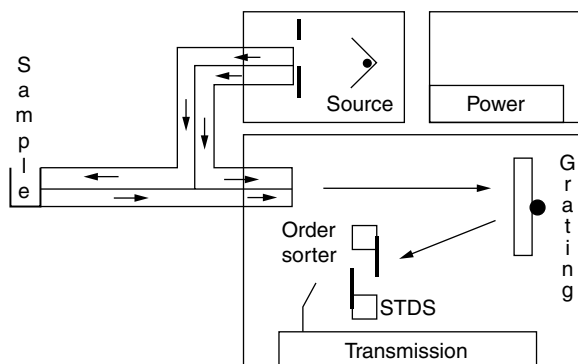


FIGURE 4.8 Diagram of a postdispersive, fast-scanning, grating spectrophotometer configured for reflectance or transmittance. (Courtesy of NIR Systems, Inc.)

element (grating). After striking the grating the light is separated into the various wavelengths prior to striking the detector(s).

Future designed systems include diode array detector advances [3], NIR-ED sources [4], Hadamard mask exit slits [5], AOTFs [6], ultrafast-spinning interference filter wheels [7], interferometers with no moving parts [8], and tunable laser sources.

4.3 COMMERCIAL INSTRUMENTATION

Table 4.3 highlights the various instrument companies, types of instrument, and wavelength coverages of the models manufactured. The type designation indicates whether the instrument is considered by the manufacturer and the industry to be a dedicated NIR instrument or whether the instrument is optimized for a different wavelength region. Although many instruments have capabilities of measuring in the NIR region, only instruments specifically optimized to the NIR region are useful for serious quantitative work. Of the dedicated NIR units, several are specifically designed for the grain or flour industry. Addresses for these companies can be found in Reference 9. The wavelength ranges and technologies stated in the table are based on information found in instrumentation reviews, such as Reference 10 and Reference 11, as well as literature provided by the manufacturers.

TABLE 4.3
Common Commercial UV-Vis-NIR, NIR, and MID-IR Instrumentation

Manufacturer	Instrument type	Type designation	Wavelength range (nm)
ABB Bomem Inc.	FT-NIR	NIR-MIR-IR	714–2632
Acton Research Corp.			
Air Instruments & Measurements			
Alliance Inst.			
American Holographic	DA	Vis-NIR	200–1100
Analect Instr (Div. of Orbital)	FT-NIR		
Analytical Spectral Devices	DA	Vis-NIR	350–2500
Analytichem Corp.			
Axiom Analytical, Inc.			
Axsun Technologies	MEMS-NIR	NIR	1310–1800
ASDI	DA/SM	Vis-NIR	350–750/1100–2500
Bio-Rad/Digilab	FT-NIR	NIR-MIR-IR	750–3000
Bomen	FT-NIR		
Bran + Luebbe	F	NIR	1445–2348
	SM	NIR	600–2500
	FT-NIR	NIR	1000–2500
	AOTS	NIR	900–1700
Brimrose	AOTF	NIR	1100–2500
			850–1700
Bruker	FT-NIR	NIR-MIR-IR	750–3000
Bruins Instruments	DM	UV-Vis-NIR	190–3000
	FTIR	NIR-MIR-IR	750–22,000
	FTIR	NIR-MIR-IR	2000–28,000
	FTIR	NIR-MIR-IR	2000–50,000
Buchi Analytical			
Buhler	FT-NIR	NIR	750–3000
Butler			

(Continued)

TABLE 4.3
(Continued)

Manufacturer	Instrument type	Type designation	Wavelength range (nm)
Carl Zeiss Jena GMBH	DM	UV-Vis-NIR	190–3100
Chemicon Inc.			
Corndisco Laboratory & Sci Grp.			
Control Development Inc.			
CVI			
Dickey-John, Churchill	F	NIR	1445–3210
DSquared Development Inc.	CCD & PbS array		
DYNA Labs, Inc.			
EG & G Princeton Applied Res.	DA	NIR	800–1700
	AOTF	NIR	1000–1700
	SM	NIR	750–2600
Elsag Bailey Applied Automation	FT-NIR		
EOS Corp.			
Equitech			
Factory of L.I. (Hungary)	F	NIR	NIT (NP)
FOSS	SM	Vis-NIR	570–1850
	SM	Vis-NIR	400–1100
	SM	NIR	800–1650
	SM	NIR	1100–2500
	SM	Vis-NIR	400–2500
	SM	NIR	800–2200
Futurex	LED	NIR	200–1100
Galileo Electro-Optics Corp.			
General Analysis Corp.	F	NIR-IR	1000–20,000
Graseby Specac			
Guided Wave	SM	NIR	800–2200
	DA	UV-Vis	230–850
	Photometer	Vis-NIR	350–2200
Hitachi Instruments, Inc.	DM	UV-Vis-NIR	187–3200
Infrared Engineering Inc.			
Infrared Fiber Systems Inc.			1100–2500
International Equipment Trading			
ISC BioExpress			
Jasco	Grating		
Katrina, Inc.	NIR-ED	NIR	900–1050
KES Analysis	DA	NIR	400–1700
			950–1700
Keft			
Labsphere			
L. T. Industries	SM	Vis-NIR	400–2400
LECO Corp.	Filter		
Leybold Inficon Inc.			
Mattson Instruments, Inc.	FTIR	MIR-IR	NIR-IR (NP)
MIDAC Corp.	FTIR	MIR-IR	2000–25,000
Moisture Systems Corp.	F	NIR	NIR (NP)
Minarad Spectrum Inc.			

TABLE 4.3
(Continued)

Manufacturer	Instrument type	Type designation	Wavelength range (nm)
NDC Infrared Engineering	F	NIR	NIT (NP)
Net Radar Inc.			
Nicolet Instruments Corp. (now Thermo Electron Corp.)			
NSG Precision Cell, Inc.			
Ocean Optics	DA	UV-Vis-NIR	200–2500
OKTec			
OLIS: On-Line Instrument Syst. (ten models available)	RSM, CD, PMT	UV-Vis-NIR	185–2600
Optical Solutions Inc. (merged with Guided Wave)	DA	UV-Vis	
Optometrics USA Inc.			
Optronic Laboratories	Grating		
Orbital Sciences Corp.	DA	NIR	800–1100
Oriel Instruments			
Oxford Analytical Inc.	F	NIR	NIR (NP)
Prinacn Analyso	(NP)	NIR	NIR (NP)
Percon GmbH	F	NIR	NIR (NP)
Perkin-Elmer Corp.	DM	UV-Vis-NIR	175–3200
	FTIR	NIR-IR	750–22,000
	FT-NIR	NIR-IR	1282–4545
Perten	DA		
Polytec Pl, Inc.	F		
Process Sensors Corp.	F (?)		
Reedy			
Reflex Analytical Corp.			
Rosemount Analytical	AOTF		
Scientific Apparatus Exchange			
SENTECH Instruments, GmbH			
Shimadzu Scientific Instruments	DM	UV-Vis-NIR	185–3200
	FTIR	NIR-MIR	800–3000
Spectral Dimensions	FPA-NIR	NIR	950–1720
Spectral Instruments, Inc.			
SpectrAlliance	DA	Vis-NIR	200–1100
Speim Grating			
Tecator			
Teledyne Brown Eng. Analyt. Inst.	Solid state		
ThIS Analytical			
Thermo Electron Corp.	FTIR		
(Thermo-Nicolet)	FT-NIR	833–2631	
(Thermo-SpectraTech)	MEMS	NIR	1350–1800
Trebor Industries	NIR-ED	NIR	NIR (NP)
	SM	Vis-NIR	
	F		Vis-NIR
Unity Scientific, Inc.	NIR	Refl. & Trans.	700–2400
Varian Optical Spectroscopy Instr.	DM	UV-Vis-NIR	185–3152
	FTIR		
Visionex Inc.			
VTT			
Wilks Enterprise, Inc.	VFA	NIR-MIR	2000–11,000
Zeltex, Inc.	DA		

It is essential that the user of NIR instrumentation have the option of presenting samples to the spectrophotometer in the most reproducible manner. Other chapters in this book describe sampling devices available for a multiplicity of applications. Instrument manufacturers are continuously developing new presentation devices for contact and noncontact reflectance; online, at-line, and side-stream monitoring; transmittance, transfectance, interactance, attenuated total reflection, and microsampling. A good general text on instrumentation is found in Reference 12.

REFERENCES

1. K. B. Whetsel, *Appl. Spectrosc. Rev.*, 2: 1–67 (1968).
2. A. Butler, *Cereal Foods World*, 28: 238 (1983).
3. Promotional information from KES Analysis Inc., New York, NY.
4. Literature from Katrina, Inc., Gaithersburg, MD.
5. Information from D.O.M. Associates, Inc., Manhattan, KS, and NIR Systems, Inc., Silver Spring, MD.
6. Promotional literature from Bran & Luebbe Analyzing Technologies, Inc., Elmsford, NY.
7. Information from Infrared Engineering, Inc., Waltham, MA.
8. Information from Hitachi Instruments, Inc., San Jose, CA (refers to UV-Vis interferometer).
9. Exhibitor Listing in Final Program, *Pittsburgh Conference*, Chicago, 1991.
10. E. W. Ciurczak, *Spectroscopy*, 6: 12 (1991).
11. Instrumentation 1991 Review, *Chem. Eng. News*, March 18, pp. 20–63 (1991).
12. R. D. Braun, *Introduction to Instrumental Analysis*, McGraw-Hill, New York, NY 1987.

5 Fourier Transform Spectrophotometers in the Near-Infrared

William J. McCarthy and Gabor J. Kemeny

CONTENTS

5.1	Introduction	79
5.1.1	FT-NIR Spectrophotometer Design.....	80
5.1.1.1	Source	81
5.1.1.2	Interferometer	81
5.1.1.3	Beamsplitter	81
5.1.1.4	Laser	81
5.1.1.5	Detector	82
5.1.1.6	Optical Components	82
5.1.1.7	Other Optical Components	82
5.2	Fourier Transform Interferometry.....	82
5.2.1	Advantages of FT-NIR Spectrophotometers.....	83
5.3	Requirements of Quantitative NIR Measurements	84
5.4	Standards	88
5.4.1	Abscissa	90
5.5	Conclusion	91
	References	91

5.1 INTRODUCTION

Near-infrared (NIR) spectroscopy has an established pedigree. When Herschel studied NIR energy in 1800, it was the first region of the electromagnetic spectrum to be discovered (after the visible region). Almost a hundred years later the Michelson interferometer is often given as the historical beginning of Fourier Transform (FT) spectroscopy. Almost a hundred years ago William Coblentz measured infrared (IR) spectra of organics. Almost 50 years ago, Karl Norris used NIR to characterize agricultural products using the new field of chemometrics. Over 25 years ago, the advent of minicomputers allowed Fourier Transform Infrared (FTIR) spectroscopy to dominate IR identification, but the chemometrics techniques were slow to be adopted in FTIR. One reason is that many spectra are required for chemometrics and a FTIR spectrum has thousands of data points compared to the hundreds of spectral data points in a dispersive NIR spectrum. The arrival of increasingly powerful computers has renewed interest in chemometrics for FT spectrometers and allowed the Fourier Transform Near-Infrared (FT-NIR) spectrophotometer industry to grow. As of today there are many companies producing FT-NIR spectrophotometers to exploit the inherent advantages of the NIR region: Analect/Orbital, Bio-Rad, Bomem, Bran + Luebbe, Bruker, Buehler, Nicolet, and Perkin-Elmer. The FT-NIR spectrophotometer shown in Figure 5.1 displays many of the common

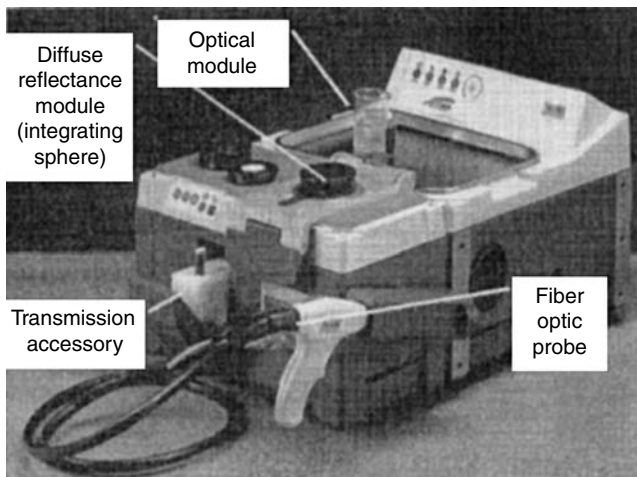


FIGURE 5.1 Typical FT-NIR analyzer showing common sampling options.

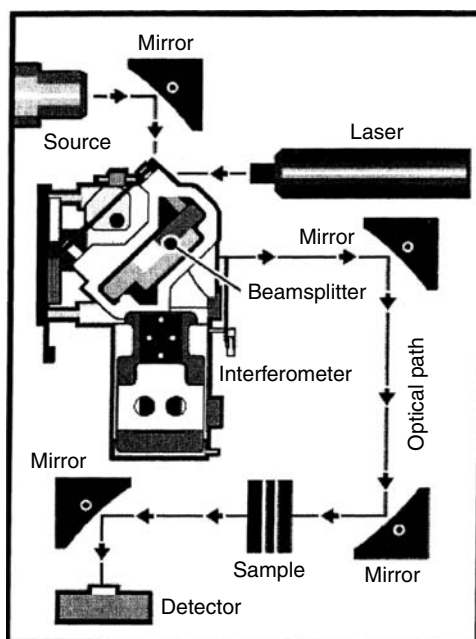


FIGURE 5.2 Optical layout of an FT-NIR spectrophotometer.

sampling options used in the NIR. Integrating spheres or other types of diffuse reflectance accessories are utilized heavily in the NIR. Fiber-optic probes can measure remote samples in diffuse reflectance or transmission. Transmission accessories or cells are often temperature controlled or heated.

5.1.1 FT-NIR SPECTROPHOTOMETER DESIGN

An FT-NIR spectrophotometer utilizes an interferometer to modulate the NIR signal and a computer (typically a PC) to obtain the spectra of materials. A simplified diagram of an FT-NIR spectrophotometer is depicted in Figure 5.2. A detailed description of an FT system has been described elsewhere [1,2].

5.1.1.1 Source

The NIR energy is supplied by the source. FT-NIR spectrophotometers typically use halogen bulbs with a wattage of 5 to 50 W. Because of the throughput advantage described in the following text, halogen bulbs are often derated (i.e., operated with reduced voltage or current) to extend their lifetime.

5.1.1.2 Interferometer

The energy of the source is collimated and directed into an interferometer where it is modulated. Each wavelength has a distinctive modulation frequency in the range of kHz, which can then be transformed by the computer into its actual electromagnetic frequency. The simplest form of an interferometer consists of two mutually perpendicular plane mirrors, one of which moves along its axis at a common velocity, and a beamsplitter.

5.1.1.3 Beamsplitter

The beamsplitter partially reflects the NIR energy onto the fixed mirror and transmits the remaining to the moving mirror. The reflected beams are recombined at the beamsplitter and directed out. The beamsplitters are layered coatings placed on a substrate. The single-beam instrument responses shown in Figure 5.3 were collected using the same detector, optics, and source so the differences observed are due to the beamsplitter. FT-NIR spectrophotometers commonly use three different types of substrates: Quartz, CaF_2 , and KBr with varying proprietary coatings.

The resulting interference between the beams depends on the optical path difference or retardation. When the fixed and moving mirror are equidistant from the beamsplitter, the retardation is zero. Neglecting losses, all the source energy reaches the detector at this point. The variation in intensity as the moving mirror is translated contains the spectral information retrieved by the FT.

5.1.1.4 Laser

FT-NIR spectrophotometer uses a monochromatic source (typically a HeNe (helium neon) laser) to control the movement of the mirror, ensure alignment of the interferometer, and provide wavelength precision. The detector signal intensity is a cosine function $I(t) = \Gamma \cos(4\pi\nu\alpha t)$ where α is the

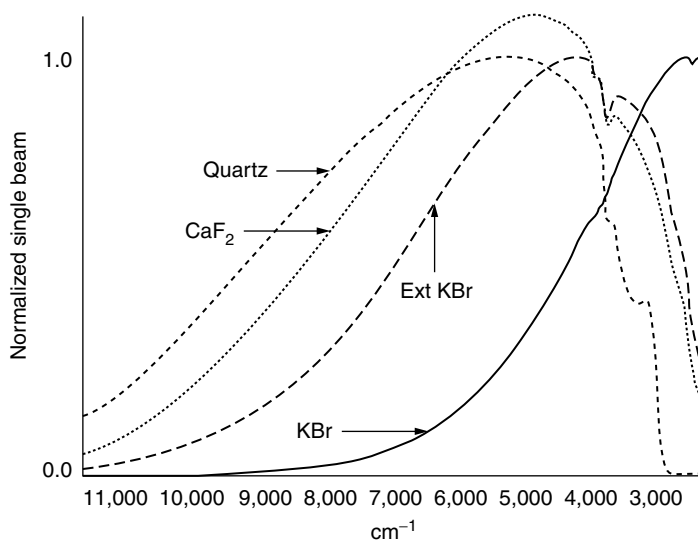


FIGURE 5.3 Single-beam curves of FT-NIR spectrophotometer beamsplitters.

TABLE 5.1
FT-NIR Detectors

Detector	Responsivity	Detectivity	Note
DTGS	Slow	Low	Flat response but relatively insensitive and can be damaged by UV
PbSe	Intermediate	Intermediate	Extremely nonlinear
PbS	Slow	Intermediate	Response is inversely proportional to modulation frequency
InSb	Fast	Very high	Must be liquid nitrogen cooled
MCT	Fast	Very high	Must be cooled (either liquid nitrogen or TE cooled)
InGaAs (2.6 micron cutoff)	Fast	High	Can also be TE cooled. Other formulations are available
InGaAs (1.7 micron cutoff)	Fast/intermediate	Very high	Gain amplification can require the detector to be run at a slower speed

optical frequency of the HeNe (15802.78 cm^{-1}) and the mirror velocity v is measured in cm s^{-1} . A typical mirror velocity for a FT-NIR spectrophotometer is $v = 0.633 \text{ cm s}^{-1}$. The HeNe laser frequency is then modulated at 20 kHz, which puts the lower frequencies in the audio range where low-noise electronics are available to digitize and transform the signal. The detector response is measured at every zero-crossing of the HeNe laser signal.

5.1.1.5 Detector

The NIR signal is delivered to the sample, which selectively absorbs and reflects or transmits the remainder to the detector. Several different detector types are available for FT-NIR spectrophotometers with advantages and disadvantages (see Table 5.1). Because FT-NIR spectrophotometers typically scan rapidly, detectors that can respond quickly are used.

5.1.1.6 Optical Components

Other optical components are needed to focus, collimate, or redirect the NIR energy. Redirection is typically accomplished using flat front surface mirrors. Aluminum is typically used because of its relatively low cost. Gold can be used and has higher reflectance in the NIR but drops off as it gets closer to the visible. Focusing or collimating optics are either mirrors or lenses. Parabolic, ellipsoidal, or aspheric mirrors are used. The mirrors are often front surfaced, either aluminum or gold. Lenses may also be used as optical elements and are typically CaF_2 , quartz, or borosilicate glass.

5.1.1.7 Other Optical Components

Other optional components may be attenuators that reduce the energy for highly transmitting samples. Wheels or flippers that contain calibration or qualification standards have also been placed internally in the FT-NIR spectrophotometer. Apertures, either adjustable irises or wheels with different size apertures, have been used to restrict the angle of the optical beam to achieve optimal resolution. Some systems have also used fixed apertures to eliminate variation.

5.2 FOURIER TRANSFORM INTERFEROMETRY

The principle of interferometry is what distinguishes FT-NIR spectrophotometers from most other spectroscopic instruments. The interference between two beams with an optical path difference or retardation creates the interferogram. A Michelson interferometer, the simplest form of an

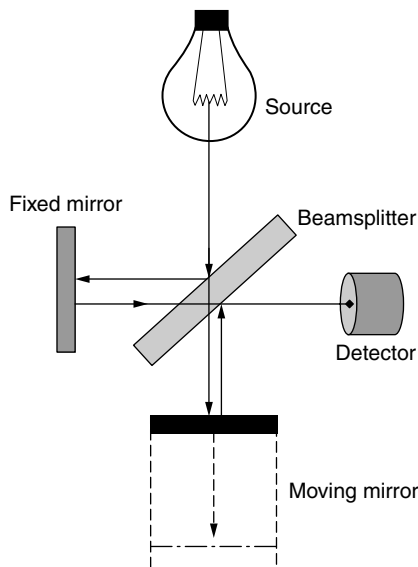


FIGURE 5.4 A Michelson (or Twyman-Green) interferometer.

interferometer, consists of two mutually perpendicular plane mirrors, one of which moves along its axis at a common velocity, and a beamsplitter (Figure 5.4). The beamsplitter partially reflects the source energy to the fixed mirror and reflects the remaining energy to the moving mirror. The beams reflected from the mirror are recombined by the beamsplitter and directed to the detector. (There are slight reflective losses, which effect the efficiency of an interferometer, but are neglected for this discussion.)

In a Michelson interferometer the optical retardation is equal to twice the moving mirror displacement. When the fixed and moving mirror is equidistant, the retardation is zero for all frequencies and the two beams interfere constructively. Therefore, neglecting losses, all the source energy reaches the detector at this point. The variation in intensity as the moving mirror is displaced contains the spectral information that is retrieved by a FT. The detector signal intensity for a broadband source is $I(t) = \Gamma(\nu) \cos(4\pi \nu vt)$. The mirror velocity ν of interferometers is generally chosen so that the modulation frequency ($2\nu v$) is in the audio range. For example, a typical mirror velocity for a rapid scanning interferometer is $\nu = 0.64 \text{ cm s}^{-1}$. If the interferometer covers the spectral range of 10,000 to 4,000 cm^{-1} , then the modulation frequencies the detector must respond to lie in the range of 12.8 to 5.1 kHz.

The interferogram of a broadband source is represented by $I(t) = \int_{-\infty}^{+\infty} \Gamma(\nu) \cos(4\pi \nu vt) d\nu$. The spectral response $F(\nu)$ of the interferometer can be calculated using the cosine FT pair of the preceding equation: $\Gamma(\nu) = \int_{-\infty}^{+\infty} I(\tau) \cos(4\pi \nu v\tau) d\tau$.

5.2.1 ADVANTAGES OF FT-NIR SPECTROPHOTOMETERS

In principle, an interferometer-based spectrometer has several basic advantages over a classical dispersive instrument:

1. *Multiplex advantage (Fellgett advantage).* In a dispersive instrument, a wavelength is measured, the grating is moved, and another wavelength is measured sequentially. If a scan takes time T , and m spatial elements are examined, the wavelength element is examined for $\Delta t = T/m$. The more spatial elements (higher resolution), the smaller amount of time that the wavelength is measured. Therefore in a dispersive instrument, assuming the main source of noise is the detector, signal-to-noise is reduced by \sqrt{m} . All frequencies in the spectra are measured simultaneously in

an FT-NIR spectrometer for the entire time T . This is because an interferometer can modulate at frequencies that are proportional to the wavelength. The time advantage is even larger, since it is directly proportional to m . A complete spectrum can be collected very rapidly and many scans can be averaged in the time taken for a single scan of a dispersive spectrometer.

2. *Throughput advantage (Jacquinot advantage)*. For the same resolution, the energy throughput in an FT-NIR spectrometer can be higher than in a dispersive spectrometer, where it is restricted by the slits. In combination with the multiplex advantage, this leads to one of the most important features of a FTIR spectrometer: the ability to achieve the same signal-to-noise ratio as a dispersive instrument in a much shorter time. The theoretical étendue or throughput of an optical system is dependent on the solid angle of the optical path Ω and the areas A of the detector and sources (or $A_s\Omega_s = A_d\Omega_d$). For an FT-NIR spectrophotometer, the solid angle is limited by the aperture (also known as the J-stop) that gives the desired resolution, $\Delta\nu$. The $\Omega_{\max} = 2\pi(\Delta\nu)/(\nu_{\max})$ for an interferometer. For a grating system the max throughput and solid angle is related to a term that relates to the focal length and the characteristics of the grating (α). The other important term is the size of the slit that gives the desired resolution $\Omega_{\text{disp}} = 2\pi\alpha(w\Delta\nu)/(\nu^2)$ where w is related to the slit width. The ratio of the solid angles gives an expression of the Jacquinot advantage $J = (\Omega_{\max})/(\Omega_{\text{disp}}) = (\nu^2)/(\alpha w\nu_{\max})$. For higher wavenumbers the Jacquinot advantage becomes greater.

3. *Connes advantage*. The intrinsic wavelength scale in an FT-NIR spectrometer provides wavelength repeatability better than one part in a million. The wave number scale of an FT-NIR spectrometer is derived from a HeNe laser that acts as an internal reference for each scan. The wave number of this laser is known very accurately and is very stable. As a result, the wave number calibration of interferometers is much more precise. If a calibration standard such as the NIST standard reference material SRM1920 is used, the wave number calibration is more accurate and has much better long-term stability than the calibration of dispersive instruments.

4. *Negligible stray light*. Because of the way in which the FT-NIR spectrometer modulates each source wavelength, there is no direct equivalent of the stray light effects found in dispersive spectrometers.

5. *Constant resolution*. In dispersive instruments, throughput is typically optimized by adjusting the slit width during the scan. Thus, signal-to-noise is constant but resolution varies. Instead, in FT-NIR spectrophotometers the resolution is defined by the J-stop (Jacquinot stop) aperture size, which does not change during data collection. Increasing the length of the scan for an interferometer increases resolution. Narrowing the slit width in a dispersive spectrometer increases resolution, but this is limited by how narrow a slit can be reliably closed. Because of this, FT-NIR spectrophotometers typically have a maximum resolution value much higher than even research-grade dispersive spectrometers.

6. *Continuous spectra*. Because there are no grating or filter changes, there are no discontinuities in the spectrum.

5.3 REQUIREMENTS OF QUANTITATIVE NIR MEASUREMENTS

One of the major obvious differences between dispersive and FT-NIR spectrophotometers is the fact that FT instruments can achieve high optical resolution without compromising signal-to-noise ratio (Figure 5.5 and Figure 5.6). An inherent advantage of the FT is that the resolution does not directly depend on mechanical light-limiting devices such as a slit. In a dispersive instrument each resolution element is a defined sliver of the total available light. The levels of signal-to-noise on the dispersive instrument detectors are limited by the light level thus the higher resolution carries the price of worse signal-to-noise ratio. High optical resolution with high signal-to-noise is important for quantitative analysis.

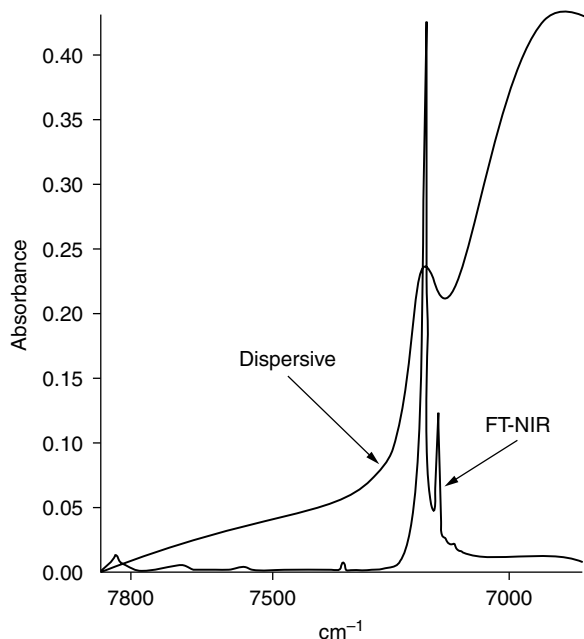


FIGURE 5.5 Comparison of spectra from a dispersive and FT-NIR spectrophotometer.

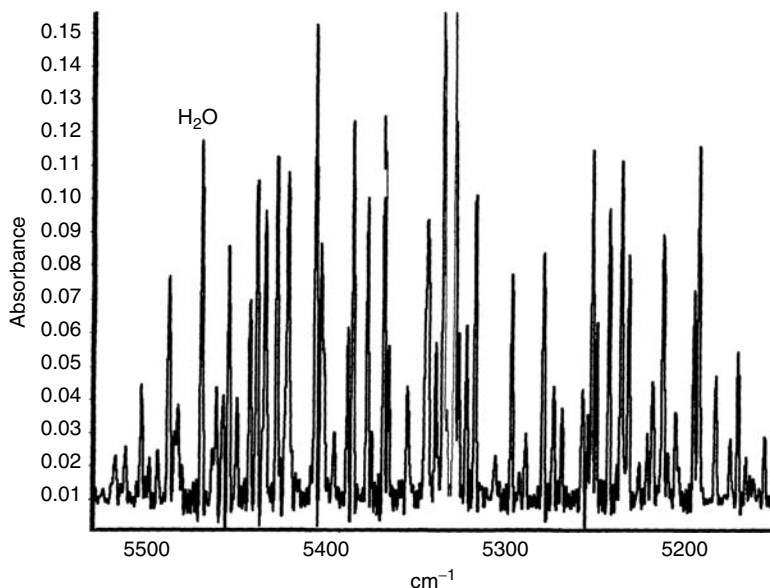


FIGURE 5.6 High-resolution water-vapor lines (first overtone).

The FT instrumentation has the proportional advantage in the mid-IR, where the whole experiment is almost always light limited. In the NIR it was argued in the past that there is just so much light that the advantages of the FT cannot be realized. It is true that the modulated and collected radiation from a quartz halogen source (10 to 50 W) can overwhelm NIR detector, heating the detector and causing nonlinearities. In actual sampling situations however, the light levels are reduced drastically, a pharmaceutical tablet for example can reduce the light levels by a factor of a million, so the light-starved detector performs much better in an FT, due to the multiplex advantage.

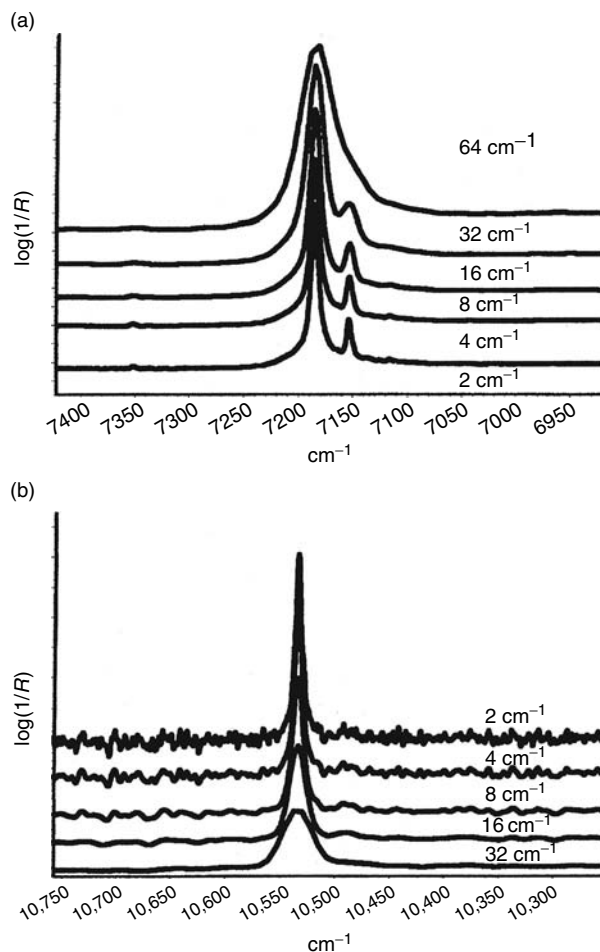


FIGURE 5.7 NIR spectrum of talc: (a) effect of resolution on band shape, (b) effect of resolution on signal-to-noise.

When evaluating FT-NIR it is important to consider real applications for the comparisons. It is obvious that much higher resolution can be achieved using an FT instrument. The higher resolution allows, for example, the ability to check the wavelength calibration using the vibrational-rotational water vapor lines [3]. Water-vapor lines in a real application could be a source of error if not completely resolved, because of the changing humidity levels in the laboratory environment. Also, many materials, have very sharp bands, which would not be resolved at 10 to 20 nm resolution of typical dispersive instruments. The better-resolved sharper bands give better chances for a selective analysis because of the increased peak height and the more characteristic spectral shape. Figure 5.7 shows a portion of the NIR spectrum of talc, where the smaller side lobe is not even resolved as a shoulder with 32 cm^{-1} resolution (equivalent to 6 nm at 7200 cm^{-1}). There are many chemicals, for example liquid hydrocarbons, some pharmaceutical materials, vitamins, and others, where the improved resolution does improve quantitative results. As is evident from Figure 5.7b, the improved resolution also reduces the signal-to-noise ratio. This is a minor peak with low response collected with one scan ($\sim(1/2)$ s) to show the trade-off of resolution and signal-to-noise in a FT-NIR spectrophotometer. The top trace shows the sharpest feature, but it is also the noisiest. In order to increase resolution, FT-NIR spectrophotometer instruments move the scanning mirror longer, thus allowing less signal averaging per unit time. In addition, the higher resolution results in more data points; thus the wavelength element is better defined. The resolution advantage of

FT allows scanning with the lowest resolution in many real applications thus collecting the highest signal-to-noise spectrum, while the optical resolution is still far superior to dispersive instruments.

A vitamin mixture was also prepared to study the effect of resolution on quantification. The actual concentrations of component were: L-ascorbic acid (vitamin C) with target value of 250 mg (varied in the calibration set from 2 to 55% of total); thiamine (B_1) target 100 mg (varied 9–30% of total); nicotinamide (niacin) 20 mg target (2–6% of total); riboflavin (B_2) 30 mg (3–7% of total); pyridoxine (B_6) target 5 mg (0.2–2% of total); filler, added to bring total weight to 500 mg (0.5–60% of total); and cellulose and hydroxypropyl methyl cellulose. The spectra were collected in diffuse reflectance mode using an integrating sphere. For most of the components, higher resolution did not show any improvement in the quantification (as measured by comparing correlation coefficients). The minor component pyridoxine did show an improvement (Figure 5.8). Resolution effects can also help in transmission experiments. For example, determining an active compared to a placebo (Figure 5.9). The multiplex advantage of the FT-NIR spectrophotometer can be readily seen in energy-limited experiments such as transmission tablet analysis. A typical layout for this experiment is seen in Figure 5.10.

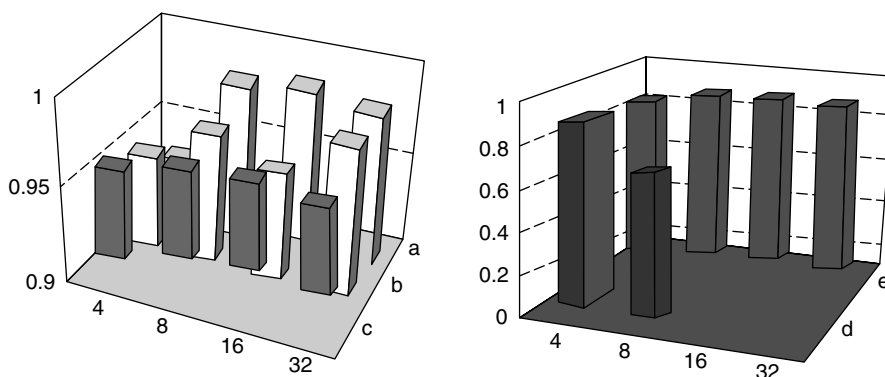


FIGURE 5.8 Correlation coefficient R^2 as a function of resolution for L-ascorbic acid a, thiamine b, nicotinamide c, riboflavin d, and pyridoxine e.

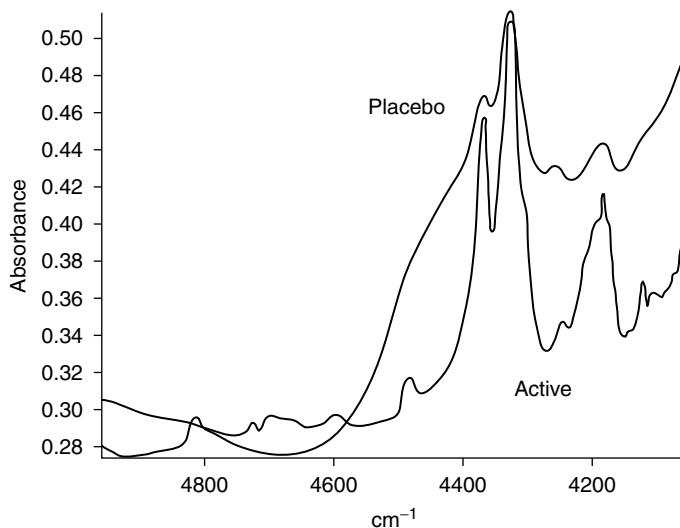


FIGURE 5.9 Spectra of an active pharmaceutical formulation compared to a placebo.

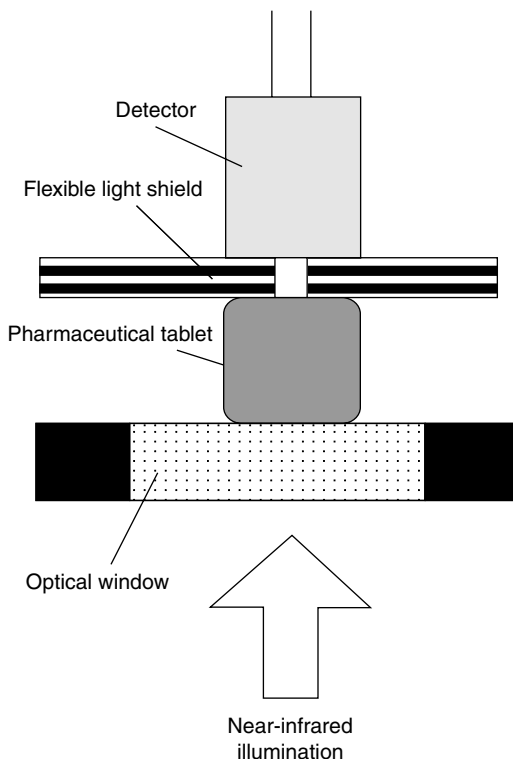


FIGURE 5.10 Schematic of transmission experiment for tablet analysis.

A typical example of a sample that requires transmission is soft-gel pharmaceutical formulations. For example, a vitamin E soft gel gives very little spectral information in reflection because of the thickness of the gelatin coating. Transmission through the vitamin shows detailed spectral information. Comparing this spectrum with the spectra of the oil extracted from the sample and measured using a fiber-optic dip probe shows the power of this technique (Figure 5.11). Studies have also been performed looking at changes of soft-gel formulation with moisture (Figure 5.12).

5.4 STANDARDS

The key assumption of most chemometric methods is that all of the spectral data have the same x -axis alignment. Even slight shifts in peak positions among spectra can create a significant cause of variance that may drastically reduce the reliability of the method. These issues are particularly important to the pharmaceutical industry. While the wavelength precision must be validated on any instrument, the requirement for consistent wavelength registration across a group of instruments is a major concern for many NIR analytical methods [4,5].

Because NIR is typically a low-resolution technique, the importance of accurate wavelengths or wavenumbers was not emphasized. The proposed USP method for wavelength accuracy defines three peaks with tolerances of ± 1 nm at 2000, 1600, 1200 nm. This translates to allowable errors of 5.0, 7.8, and 13.9 cm^{-1} . A typical FT-NIR spectrophotometer has a precision of better than 0.01 cm^{-1} . In Figure 5.13, a three-week monitoring of the water-vapor peak at 7299 shows a standard deviation of 0.0027 cm^{-1} (0.0005 nm). This precision has required careful evaluations of the standards that are used for NIR.

However, even in FT-NIR spectroscopy the actual wavelength accuracy depends to a small degree on the optical design of the instrument. The measured wavenumber $\nu' = \nu\{1 - (1/2)\alpha^2\}$ where α is

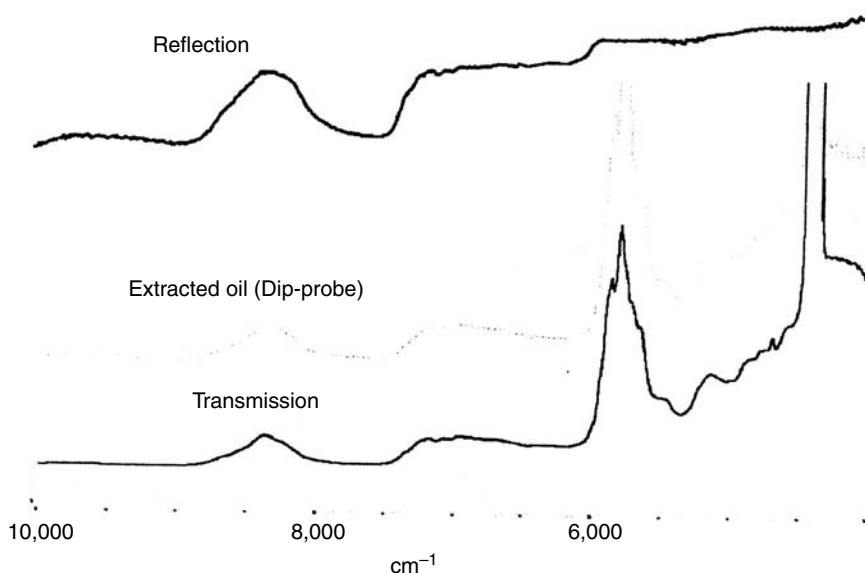


FIGURE 5.11 Comparing vitamin E spectra of reflection, transmission, and extracted oil from inside the capsule.

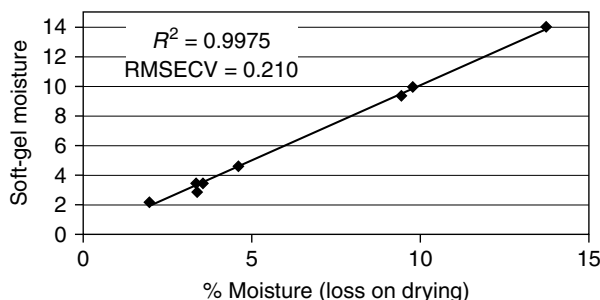


FIGURE 5.12 NIR-derived moisture values compared to actual values.

related to the etendue or aperture. The differences are a linear shift and the observed differences are typically $< 0.1 \text{ cm}^{-1}$ deviation [3]. To correct for the differences, the spectral features of a known standard are measured and the laser frequency adjusted to shift the spectrum. In practice, this is rarely done because the observed differences are small.

Bomem has reported using toluene as test of both wave number and absorbance repeatability and reduced differences to less than $\pm 0.1\%$ [5].

Many FT-NIR spectrophotometers are designed to have internal standards for calibration procedures. The most common internal standard is a polystyrene sheet (Figure 5.14) used in transmission. Polystyrene has also been used in reflection.

A sheet of polystyrene was compared to SRM 1921 (the NIST mid-IR standard) and shown that a thicker sample of polystyrene plastic that has been validated with the SRM 1921 standard would make an excellent reference material for verifying wavelength accuracy in a medium resolution FT-NIR spectrophotometer [6]. Traceable standards such as the SRM 2035 and SRM 1920 \times are also used and designed to fit in the sample position.

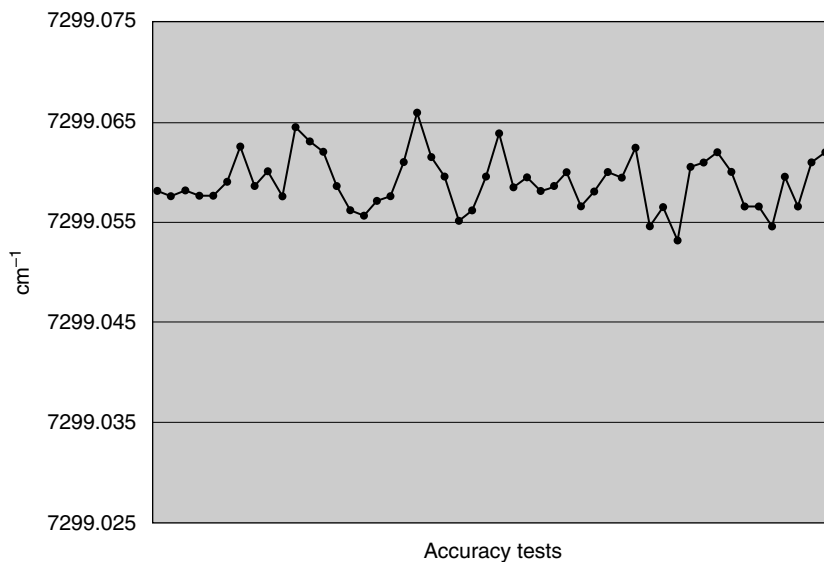


FIGURE 5.13 Trend line of position of water-vapor peak.

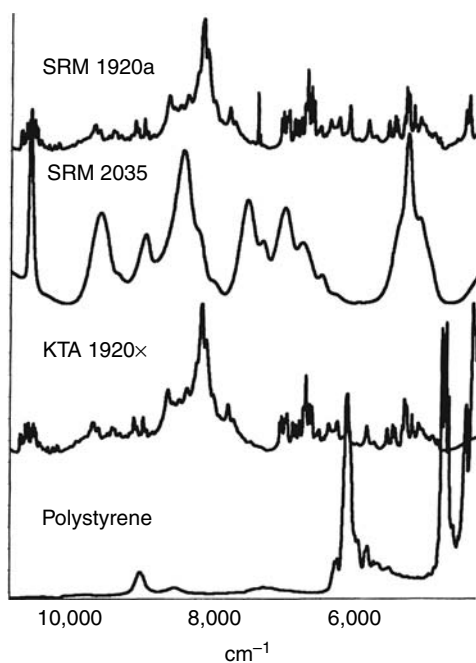


FIGURE 5.14 NIR standards.

5.4.1 ABSCISSA

Photometric linearity is an important area as well. Photometric qualification is based on a set of transmission or reflectance standards with known values. In transmission, filters with known transmittance values are used. In reflectance, Labsphere™ Spectralon gray standards with reflectance values from 0.99 to 0.02 are used.

5.5 CONCLUSION

In the last decade, FT-NIR spectrophotometers have made inroads into the traditional NIR applications areas: particularly the pharmaceutical, petrochemical, and chemical markets. The advantages of FT provide the ability to produce accurate, reproducible spectra for identification, and quantification of difficult samples. This makes FT-NIR an invaluable tool for quality control and quality assurance applications.

REFERENCES

1. P. R. Griffiths and J. A. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley, New York, 1986.
2. R. J. Bell, *Introductory Fourier Transform Spectroscopy*, Academic, New York, 1972.
3. S. R. Lowry, J. Hyatt and W. J. McCarthy, *Applied Spectroscopy* (submitted for publication).
4. United States Pharmacopeia, Near-Infrared Spectrophotometry, *Pharmacopeial Forum*, 24: 1998.
5. H. Buijs, The Promise of FT-NIR: Universal Calibrations.

6 Analysis Using Fourier Transforms

W. Fred McClure

CONTENTS

6.1	Introduction	93
6.2	Fourier Mathematics	94
6.3	The Fourier Transformation	95
6.3.1	The Periodicity Requirement	95
6.3.2	Fourier Transforms	97
6.4	Qualitative Spectral Analysis	100
6.4.1	Convolution Smoothing in Wavelength Space	100
6.4.1.1	Convolution	100
6.4.1.2	Moving Point Average Smoothing	100
6.4.1.3	Polynomial Smoothing	101
6.4.1.4	Fourier Smoothing	102
6.4.2	Deconvolution via Fourier Space	104
6.4.3	Computing Derivatives via Fourier Space	106
6.4.3.1	Fourier Derivatives	108
6.4.4	Corrects for Particle Size Anomaly	109
6.5	Quantitative Spectral Analysis	109
6.5.1	Estimating Composition with Fourier Coefficients	109
6.5.2	Cutting Computer Storage Requirements by 98%	111
6.5.3	Reducing Calibration Times by 93%	112
6.5.4	Encouraging Calibration Maintenance	112
6.5.5	Reducing Multicollinearity	113
6.6	Artificial Intelligence	113
6.6.1	Spectral Searching, Matching, and Component Identification	113
6.6.2	Checking Instrument Noise in Real Time	115
6.6.3	Testing for Instrument Anomalies in Real Time	115
6.7	Interferogram Space	117
6.7.1	Estimating Chemistry from Interferogram Data	117
6.7.2	No-Moving-Parts FT-NIR	118
	References	119
	Bibliography	119

6.1 INTRODUCTION

Fourier transform infrared (FTIR) techniques are well-established techniques [1–3] and have been for a long time. However, the use of Fourier transforms in the analysis of near-infrared (NIR) spectra

had not been tried until 1981, especially Fourier analysis of NIR spectra obtained from dispersion instruments. Our [1]* excitement over the use of Fourier transforms in the analysis of NIR data was stimulated by work that was ongoing in the field of information processing. There had been some very exciting results in using Fourier transforms in the transmission of voice information. Electrical engineers had demonstrated that voice information could be carried in just a few low-frequency Fourier coefficients. The work produced magnificent results. First, the voice information or signal was digitized. The Fourier transform was taken of the signals and the Fourier coefficients were transmitted down the line. However, only the first few Fourier coefficients were transmitted and, on the other end, the Fourier transform was taken again, giving intelligible voice information. Some fidelity, such as the distinction between male and female voice, was lost but the content of the information was there.

The driving force behind our interest in this area at that time was three problems encountered working in wavelength space. First there were too many data in a spectrum. At that time we were recording 1700 data points per spectrum from 900 to 2600 nm. Archiving spectra required a tremendous amount of magnetic storage and the problems of retrieval were becoming insurmountable. Second, the data within a spectrum were highly intercorrelated, a phenomenon referred to as multicollinearity. This meant there were superfluous data in the spectrum, but there was no sound basis for discarding any. Third, NIR calibrations based on wavelength data were rather unstable. That is to say, if new spectra were added to a calibration set and another multilinear calibration obtained, the wavelengths included in the new calibrations may not be the original ones. There were no fundamental principals in place at that time for selecting the *most* appropriate wavelengths for calibration equations.

It occurred to us that if sufficient voice information was encoded in the first few Fourier coefficients to give intelligible information when “retransformed,” we should be able to use this same technique to compress NIR spectra to a smaller set of numbers. This in turn would result in much faster and more stable calibrations. We performed our first experiments in 1978 but, because the results were so unbelievable, we did not publish the data until much later [4,5]. We were extremely pleased to discover that as few as six pairs of Fourier coefficients, just 12 numbers, from $\log(1/R)$ spectra could be used to estimate chemistry in pulverized samples of tobacco without loss of accuracy or precision [5]. Since that time we have demonstrated that a number of advantages can be accrued by working in the Fourier domain (FD) rather than the wavelength domain (WD). In this chapter, we will discuss the advantages of working in Fourier space from the qualitative (special transformations) and quantitative (chemical determinations and efficiency) standpoint.

6.2 FOURIER MATHEMATICS

Any well-behaved periodic function can be represented by a Fourier series of sine and cosine waves of varying amplitudes and harmonically related frequencies. This may be illustrated using the square wave function given in Figure 6.1a. Couching the illustration in spectroscopy jargon, let us assume that the square wave spectrum represents an absorption band with an absorbance of 1.0 centered at 1300 nm. We know that this spectrum can be defined mathematically in terms of sine and cosine terms as follows:

$$f(\lambda) = a_0 + (a_1 \sin \omega\lambda + b_1 \cos \omega\lambda) + (a_2 \sin 2\omega\lambda + b_2 \cos 2\omega\lambda) + \cdots + (a_k \sin k\omega\lambda + b_k \cos k\omega\lambda) \quad (6.1)$$

* The author gratefully acknowledges the collaboration of Dr. Abdul Hamid in the early stages of the development of Fourier analysis of NIR data.

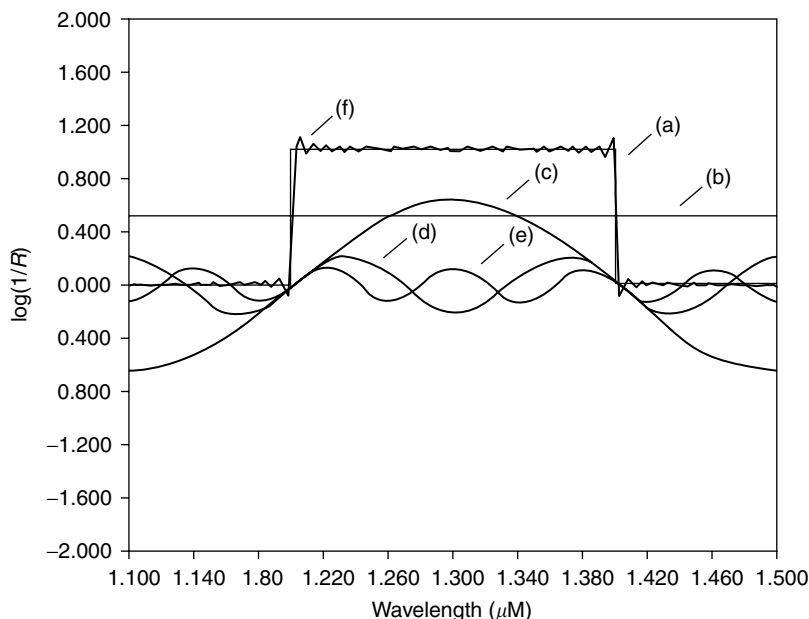


FIGURE 6.1 Square-wave spectrum (a), inverse Fourier transforms using a_0 and b_0 (b), a_1 and b_1 (c), a_2 and b_2 (d), a_3 and b_3 (e), and a_0, b_0 , through a_{49} and b_{49} (f).

The first term a_0 is the average value of the spectrum (called the *mean term*, Figure 6.1b). The weighting factors a_i, b_i are sometimes referred to as coefficient pairs. The magnitude of a_i and b_i determines the amplitudes of the contributing terms. We will show later that these are really the Fourier coefficients where $i = 1, 2, 3, \dots, k$ is the Fourier index.

For periodic functions the b_0 coefficient is always zero. The terms involving a_1 and b_1 are the sine and cosine components with one oscillation in the interval 1200 to 1300 nm (Figure 6.1c); the terms involving a_2 and b_2 are the sine and cosine components with two oscillations in this interval (Figure 6.1d); the terms involving a_3 and b_3 are the sine and cosine components with three oscillations in this interval (Figure 6.1e). Hence the terms involving a_i and b_i are the sine and cosine components having i oscillations in this interval. Including enough terms in Equation (6.1) (say, 50 pairs) will produce an excellent facsimile of the square wave (Figure 6.1f). Table 6.1 gives the coefficients for the square wave transformation; note that even-numbered coefficients are always zero, a characteristic peculiar to the square waveform.

6.3 THE FOURIER TRANSFORMATION

6.3.1 THE PERIODICITY REQUIREMENT

If a square waveform can be adequately represented by a series of sine and cosine terms, then any NIR spectrum can be represented as a Fourier series of sine and cosine waves of varying amplitude and harmonically related frequencies provided the spectrum is periodic and continuous. A typical NIR spectrum of tobacco seeds is given in Figure 6.2a. From the standpoint of periodicity, there exists a discontinuity on the right end of the spectrum due to the dominant linear trend upward as wavelength increases. That is, if we lay the same spectrum end-to-end repeatedly, this discontinuity makes the spectrum aperiodic. When we use multiple linear regression (MLR) in wavelength space [16] for the purpose of computing the chemical content of a series of samples, this linear trend as well as the discontinuity can be disregarded. However, it cannot be ignored if a spectrum is to be

TABLE 6.1
First 40 Fourier Coefficients of the Square-Wave Spectrum
in Figure 6.1a^a

Fourier index	Value of coefficient	Fourier index	Value of coefficient
a_0	400.00		
a_1	-254.64	a_{11}	23.09
b_1	2.00	b_{11}	-2.00
a_2	0.00	a_{12}	0.00
b_2	0.00	b_{12}	0.00
a_3	84.87	a_{13}	-19.52
b_3	-1.99	b_{13}	2.00
a_4	0.00	a_{14}	0.00
b_4	0.00	b_{14}	0.00
a_5	-50.90	a_{15}	16.90
b_5	2.00	b_{15}	-2.00
a_6	0.00	a_{16}	0.00
b_6	0.00	b_{16}	0.00
a_7	36.34	a_{17}	-14.89
b_7	-2.00	b_{17}	2.00
a_8	0.00	a_{18}	0.00
b_8	0.00	b_{18}	0.00
a_9	-28.25	a_{19}	13.30
b_9	2.00	b_{19}	-2.00
a_{10}	0.00	a_{20}	0.00
b_{10}	0.00	b_{20}	0.00

^a Note that even-numbered coefficients are all zero, a characteristic of the square waveform.

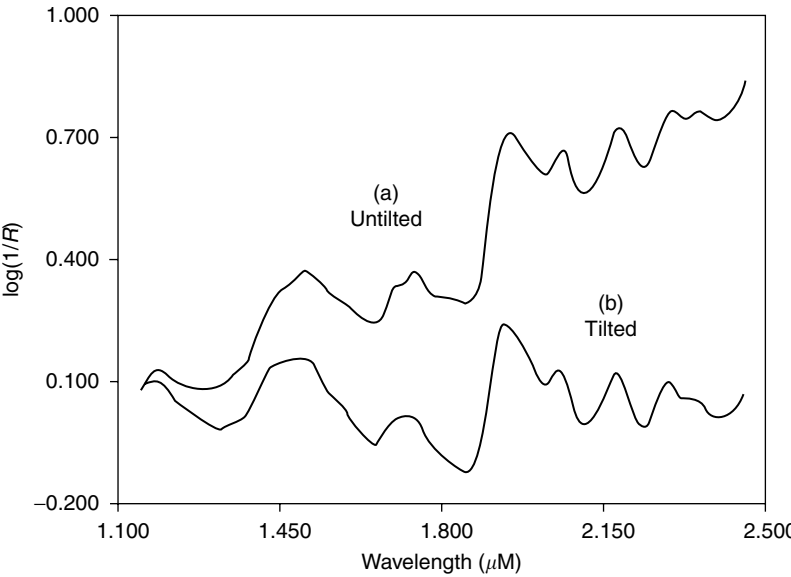


FIGURE 6.2 Untilted or original (a) and tilted (b) spectrum of wool. The tilted spectrum satisfies the periodicity requirement of the Fourier transform.

transformed to the FD for analysis. One effective method of compensating for this problem is to adjust or “tilt” the spectrum by replacing $f(\lambda_h)$ by

$$f(\lambda_h) - m \left(\frac{f(\lambda_h) - f(\lambda_1)}{f(\lambda_n) - f(\lambda_1)} \right) \quad (6.2)$$

where m is equal to the observed $f(\lambda_n) - f(\lambda_1)$. If the m value for each spectrum is stored, then the effect of the adjustment can be removed when the spectrum is recalculated. Otherwise, the Fourier transformation will generate artifacts at the beginning and the end of the recalculated spectrum. The effect of this tilting operation can be seen in Figure 6.2. Note how Figure 6.2a slopes upward so that the end points of the spectrum are at two different levels. Laying this spectrum end-to-end along the wavelength axis leaves a discontinuity at the beginning of each period. However, one can readily appreciate that laying Figure 6.2b, the tilted spectrum, end-to-end repeatedly will produce a periodic waveform with one period covering the range from 1150 to 2449 nm.

6.3.2 FOURIER TRANSFORMS

If $F(1), F(2), \dots, F(N-1), F(N)$ represents the recorded spectral values at N equally spaced wavelengths, denoted by $1, 2, 3, \dots, \lambda$, and if $N = 2M$ is even, then it is well known that the spectrum can be approximated by:

$$F(\lambda) = a_0 + \sum_{k=1}^{M-1} a_k \cos\left(\frac{2\pi k\lambda}{N}\right) + \sum_{k=1}^{M-1} b_k \sin\left(\frac{2\pi k\lambda}{N}\right) + a_M \cos(\pi\lambda) \quad (6.3)$$

for $\lambda = 1, 2, \dots, N$ (wavelength index) and $k = 1, 2, \dots, M$ (Fourier index). The values of a_k and b_k , the Fourier coefficients, are computed easily via the mixed-radix fast Fourier transform (FFT) developed for efficiently computing the following definitions:

$$a_0 = \frac{1}{N} \sum_{x=1}^N f(\lambda) \quad (6.4)$$

$$a_k = \frac{2}{N} \sum_{x=1}^N f(\lambda) \cos\left(\frac{2\pi k\lambda}{N}\right) \quad (6.5)$$

for $k = 1, 2, 3, \dots, M-1$

$$b_k = \frac{2}{N} \sum_{x=1}^N f(\lambda) \sin\left(\frac{2\pi k\lambda}{N}\right) \quad (6.6)$$

for $k = 1, 2, 3, \dots, M-1$

$$a_M = \frac{1}{N} \sum_{x=1}^N f(x) \cos(\pi\lambda) \quad (6.7)$$

One of the nice features of the Fourier approach is that k (the number of Fourier coefficients) can be kept reasonably small and yet provide an excellent facsimile of the recorded NIR spectrum [4]. Figure 6.3 illustrates this feature. Spectrum (a) in Figure 6.3 is an originally recorded spectrum of wool. Spectrum (b) is the Fourier-reconstructed spectrum using only the first 50 pairs of coefficients; it has been shifted upward to permit the reader to compare the reconstructed spectrum with the original. Figure 6.4 is the difference spectrum obtained by subtracting the original spectrum (a) from the model spectrum (b); this “noise” spectrum has a standard deviation of 230 microlog (1/R) units,

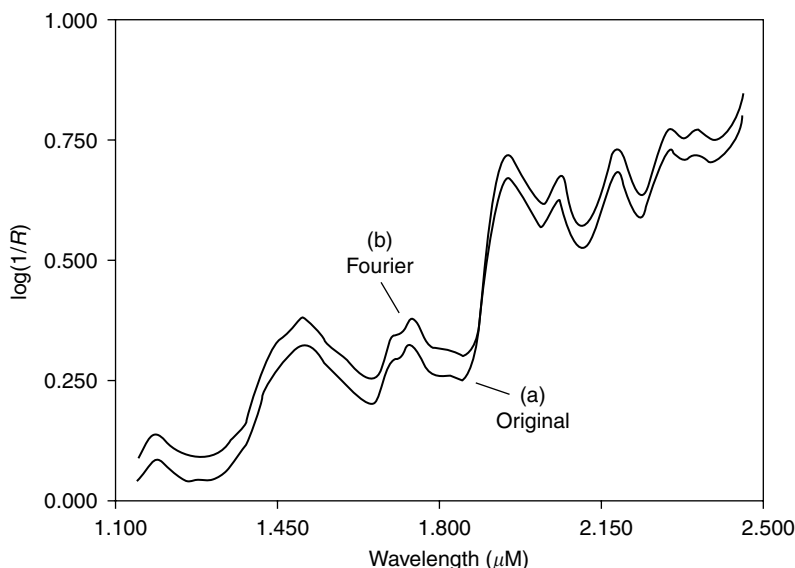


FIGURE 6.3 Original (a) and recomputed spectrum (b) from the first 50 pairs of Fourier coefficients.

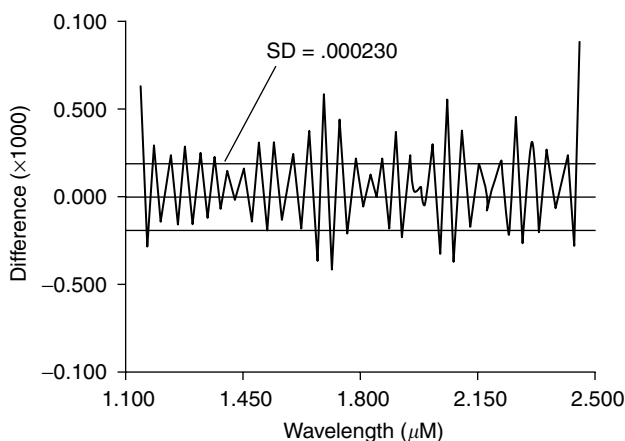


FIGURE 6.4 Difference spectrum obtained by subtracting Figure 6.3a from Figure 6.3b. Note that the standard deviation of the difference is only 230 microabsorbance units.

giving an r^2 of fit for the two spectra of .999988. This implies that most of the spectral information is contained in the first 50 pairs of Fourier coefficients.

The output from the FFT may be stored in the form of a Fourier spectrum [4] much as a wavelength spectrum is stored. The arrays can be of the form $a_0, a_1, b_1, a_2, b_2, \dots, a_{M-2}, b_{M-2}$. The b_0 term (where $\sin 0 = 0$) is dropped from the array because it makes little contribution to the spectral information and can easily be accounted for in other ways. The coefficient pairs are arranged conveniently in the Fourier spectra in order of increasing frequency. For example, the Fourier spectrum of the *original* spectrum in Figure 6.3 is shown in Figure 6.5. Table 6.2 gives the first 40 Fourier coefficients for this transformation. Since the complete Fourier spectrum is a mirror image about zero, only the right half need be stored. This arrangement makes it easy to apply apodization functions to the data for purposes of smoothing, self-deconvolution, derivatives, etc. It should be

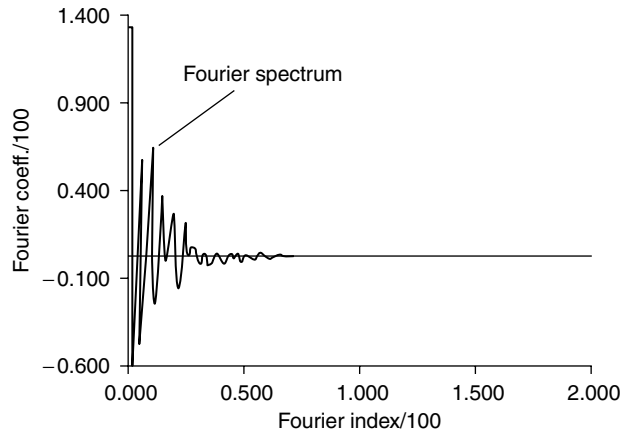


FIGURE 6.5 Fourier coefficients are stored in a Fourier spectrum typical of the one above. Note that most of the information is below the first 100 coefficients or 50 pairs. Only the right half of the Fourier spectrum is stored. Both the real a_i and the imaginary b_i coefficients are stored alternatively in the same spectrum (see Table 6.1).

TABLE 6.2
First 40 Fourier Coefficients of the Spectrum of Wool
in Figure 6.5^a

Fourier Index	Value of coefficient	Fourier index	Value of coefficient
a_0	131.94		
a_1	20.57	a_{11}	-15.13
b_1	-17.27	b_{11}	-18.54
a_2	-59.28	a_{12}	-9.37
b_2	28.27	b_{12}	19.57
a_3	55.56	a_{13}	1.88
b_3	-49.84	b_{13}	1.34
a_4	-6.72	a_{14}	6.15
b_4	-1.52	b_{14}	5.30
a_5	27.20	a_{15}	4.51
b_5	62.18	b_{15}	-0.91
a_6	-21.84	a_{16}	-4.07
b_6	-28.76	b_{16}	1.72
a_7	-2.27	a_{17}	2.94
b_7	34.74	b_{17}	-4.56
a_8	6.07	a_{18}	-4.27
b_8	-2.77	b_{18}	-2.84
a_9	4.77	a_{19}	1.11
b_9	18.08	b_{19}	2.72
a_{10}	24.71	a_{20}	1.43
b_{10}	-0.53	b_{20}	-0.57

^a Note that none of the coefficients are zero. This is a characteristic of real NIR spectra.

mentioned at this point that the Fourier transform of an N point spectrum can have as many as N points, or $N/2$ a_i values and $N/2$ b_i values.

Finally, we should also define

$$E(f) = \mathcal{F}\{\mathcal{F}(\lambda)\} \quad (6.8)$$

and

$$\mathcal{F}(\lambda) = \mathcal{F}^{-1}\{E(f)\} \quad (6.9)$$

where \mathcal{F} and \mathcal{F}^{-1} are the Fourier transform and the *inverse* Fourier transform, respectively. The word *inverse* implies direction, not a different form of the Fourier transform. The Fourier transform of a set of Fourier coefficients produces the spectrum from which the Fourier coefficients were derived.

6.4 QUALITATIVE SPECTRAL ANALYSIS

6.4.1 CONVOLUTION SMOOTHING IN WAVELENGTH SPACE

Smoothing is one of the first operations performed on recorded NIR spectra. Its purpose is to remove as much noise as possible from spectra without excessively degrading important information. The three most popular methods for smoothing spectra in wavelength space are (a) moving point average (MPA), (b) polynomial (sometimes referred to as the Savitzsky–Golay method), and (c) Fourier. The first two are performed in wavelength space; the third is performed using Fourier space. The first two methods involve “convolution” of two functions in wavelength space. Fourier smoothing achieves the same objective by a simple multiplication of two functions in Fourier space plus the Fourier transforms to and from Fourier space. So, what is convolution?

6.4.1.1 Convolution

The convolution of two functions $f(\lambda)$ and $g(k + \lambda)$ over the interval from $-M$ to M is usually written as

$$f(\lambda)^*g(\lambda) = \sum_{k=-M}^{+M} f(\lambda)g(k + \lambda) \quad (6.10)$$

Now convolution in and of itself is rather boring, but the tedium goes away when you realize that you can take Equation (6.10) and do something useful with it, such as smoothing.

6.4.1.2 Moving Point Average Smoothing

Moving point average smoothing is an easy way to understand convolution. It is explained here to dismiss the myth put forth by many writers that convolution is easier calculated than visualized or understood. Basically, smoothing computes the average of an odd number of sequential points, replacing the center point with the average value, or

$$f_c(\lambda) = \left(\frac{\sum_{i=1}^N f_i(\lambda)}{N} \right) \quad (6.11)$$

where $f_c(\lambda)$ is the calculated center point of the interval, $f_i(\lambda)$ equals the i th value of the spectral data in the interval, and N is any odd number of sequential points. The MPA smoothing process usually starts on the left end of the spectrum (at the shortest wavelength) and moves to the right one data point at a time until the right end of the spectrum is reached. Because the center point of the interval is replaced with the average, $(N - 1)/2$ points on each end are not included in the smoothed spectrum (a problem obviated by using Fourier transforms). The severity of MPA smoothing increases as N increases.

Now here's the interesting part. Note that Equation (6.11) is very similar to Equation (6.10). The only difference is that Equation (6.11) lacks the multiplication of $f(\lambda)$ by $g(k + \lambda)$ and the division by N . So let's rewrite Equation (6.11) such that

$$f_c(\lambda) = \left(\frac{\sum_{k=-M}^{+M} f(\lambda) g(k + \lambda)}{N} \right) \quad (6.12)$$

Now $f_c(\lambda)$ is still the calculated center point of the interval, but the interval is defined from $-M$ to $+M$. This interval is called the convolution interval. Also, $N = (2M + 1)$ is the total number of points in the convolution interval, and is usually an odd number of points. The spectral index λ is the point of reference for the convolution function such that when $\lambda = 0$ the convolution function $g(k)$ is positioned so that the $-M$ point is lined up with the first point in the spectrum. Each point of $g(k)$ from $-M$ to M is multiplied by the corresponding (aligned) points in the spectrum, and the values are summed and divided by N . The result replaces the spectral value corresponding to the aligned center point of the convolution function. Then λ is set to 1, which moves the "boxcar function" to the right one point and the process is repeated and so on until the boxcar function reaches the right end of the spectrum. Hence, Equation (6.12) requires that $g(k + \lambda)$ slide point by point, similar to a template, along the axis of the spectrum from $\lambda = 0$ to $(P - N)$ where P is the total number of points in the spectrum. If we require that all values of $g(k)$ be equal to 1 for all values of k from $-M$ to $+M$, $g(k)$ will look like Figure 6.6b, a boxcar-shaped function with an amplitude of 1.0 and a width of N points. Hence, you can see why MPA smoothing is referred to as boxcar smoothing. The MPA smoothing process (referring to Figure 6.6a–c) on any NIR spectrum, $f(\lambda)$ (Figure 6.6a), will produce the spectrum in Figure 6.6c, a spectrum that lacks resolution.

Since we have restricted the values of $g(k + \lambda)$ to 1, Equation (6.12) reduces nicely to Equation (6.11) where the process of multiplying $f(\lambda)$ by $g(k + \lambda)$ across the interval $-M$ to M is omitted from the notations. Hence, Equation (6.12) requires the computation of the convolution for each value of λ from 0 to $(P - M)$, where P is the number of points in the spectrum. Note also that M points must be dropped from the left end and the right end of the smoothed spectrum. Another interesting thing about convolution smoothing using Equation (6.12) is that the convolution function may take different forms (triangular, cosine, etc.) depending on the type of spectral data and the objectives of the smoothing process [6].

Convolution of an instrument slit function with an absorption band can be illustrated using Figure 6.7. Let us assume that the function in Figure 6.7b is the slit function of a scanning spectrometer having a bandpass of 9 nm. Also, let us assume that the spectrum in Figure 6.7a is the absorption band of an analyte, whose full width at half-height (FWHH) is 10 nm, to be scanned by the spectrometer. The recording process convolves the slit function with the absorption band yielding the spectrum in Figure 6.7c whose amplitude is diminished and whose FWHH is broadened. As a general rule, unless the bandpass of the spectrometer is considerably narrower than the FWHH of the absorption band, the recorded band will always have a diminished peak height and a broadened FWHH. We will show later how Fourier self-deconvolution (FSD) can be used to remove the slit function from the spectral absorption bands.

6.4.1.3 Polynomial Smoothing

Polynomial smoothing involves statistically fitting, by least-squares techniques, a polynomial of the form

$$F(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_n\lambda^n \quad (6.13)$$

to an odd number of sequential spectral data points and computing the center point of the interval from the polynomial [6]; n is the degree of the polynomial. The procedure then requires that one point on the left end of the interval be dropped and a new point on the right of the interval be

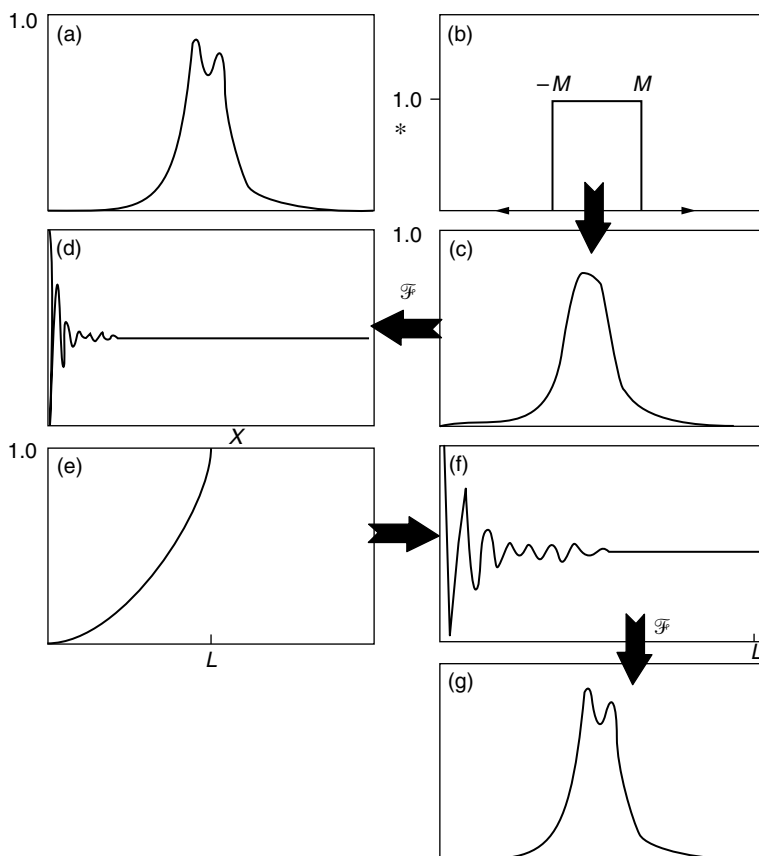


FIGURE 6.6 Illustrations of convolution and FSD: (a) A spectrum consisting of two overlapping absorption bands whose peaks are clearly resolved. (b) A boxcar convolution function. (c) MPA smoothed form of (a) having resolved peaks. (d) Fourier transform of (c). (e) Exponential apodization function, $\exp(2\beta|x|)$. (f) Apodized Fourier spectrum accenting the high-frequency coefficients. (g) The self-deconvolved spectrum showing clearly resolved peaks. Parts a, b, and g are WD representations; d, e, and f are FD.

picked up. New regression coefficients are calculated and the polynomial is used to compute the second center point. The fitting and calculating process slides to the right one point at a time until the right end of the spectrum is reached. Seldom would you use a polynomial of a degree higher than five (quintic) for NIR data. Note, although trivial, that no smoothing will be imposed on the spectral data if the number of points in the interval is one less than the degree of the polynomial. That is, a polynomial of degree n will fit $(n - 1)$ points perfectly. Of course, the polynomial method could be used to recompute all points in the interval, but this is seldom done due to the uncertainty of fit in the regression for points removed from the center point of the interval. Therefore, $(N - 1)/2$ points, where N is an odd number of points over which the polynomial is fitted, are usually dropped just as in the case of the MPA smoothing method. Polynomial smoothing, like MPA smoothing, works on one interval at a time, moving to the right one point at a time.

6.4.1.4 Fourier Smoothing

Fourier smoothing achieves the same objective as MPA and polynomial methods but without the loss of end points. The process seems rather circuitous at first, but a closer look at the results makes the process more appealing. Referring to Figure 6.8, the spectrum $f(\lambda)$ is transformed to the

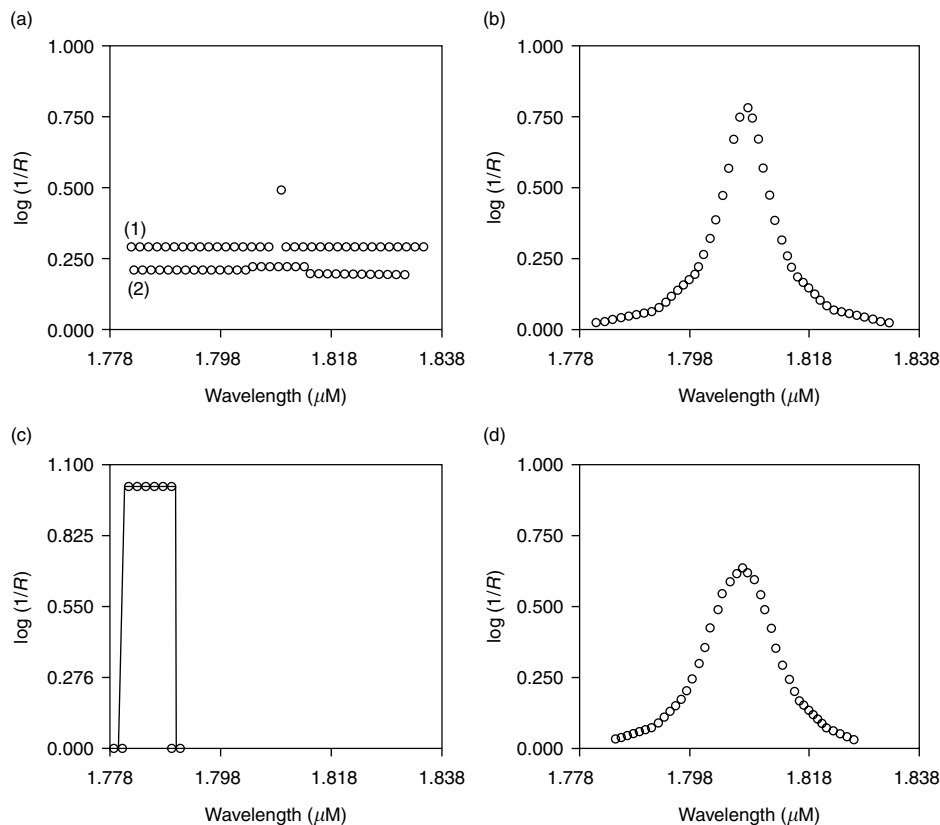


FIGURE 6.7 MPA smoothing: (a1) section of the spectrum in Figure 6.9a, (a2) smoothed by a 9-point MPA smooth, (b) a Lorentzian band with a peak height of 0.75 and FWHH of 10 nm, (c) the MPA convolving function $g(x)$, and (d) convolution of Part b.

FD giving the Fourier spectrum b1. This spectrum is then multiplied by an apodization function (a boxcar function is shown in b2). Finally, the Fourier transform of the truncated Fourier spectrum is computed, yielding the smoothed spectrum.

The effect of smoothing spectral data can be appreciated by looking at Figure 6.9 and Table 6.3. Figure 6.9a is a spectrum of wool scanned over the range 1150 to 2500 nm. One data point at 1800 nm has been changed from a recorded value of 0.323 to 0.5. $\log(1/R)$ units. This change appears as a noise spike on the original spectrum in Figure 6.9a. In the 25-point MPA smooth spectrum the spike has been diminished significantly, but in the process the values of the surrounding data points have been drastically elevated (Figure 6.7a and Figure 6.9b). In addition, 12 points on either end of the spectrum have been dropped as a result of the convolution process. On the other hand, Fourier smoothing has a much less severe effect on surrounding data and end points are not dropped from the spectrum.

The two protein bands (2055 and 2077 nm) and the water band (1941 nm) in Figure 6.9a are marked for reference later. Two things should be noted from Table 6.3. First, MPA smoothing degraded the water band from 0.7168 to 0.7115 as the convolution interval increased from 1 (no smoothing) to 25 points. At the same time the peak of the water band shifted from 1941 to 1943 due to a smoothing moment caused by the asymmetry of the water band. *Note this point:* If a band is symmetrical there will be no shift of the peak no matter how large the convolution interval. This one fact makes it impossible to determine a priori the optimum convolution interval. Consequently, unless you know a lot about the characteristics of the noise in your spectrometer plus detailed

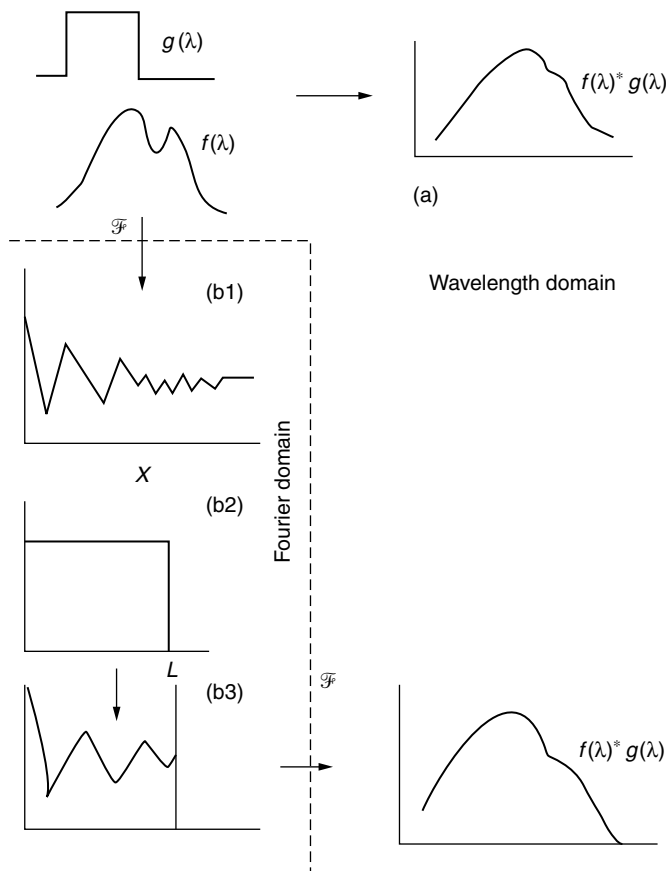


FIGURE 6.8 Convolution smoothing in the WD versus smoothing via the FD. MPA smoothing involves the convolution of two functions in wavelength space, but only multiplication of two functions in Fourier space.

information concerning the FWHH of the absorption bands, judgments concerning smoothing are purely empirical.

The effect of boxcar, polynomial, and Fourier smoothing can be seen by expanding the region around the noise spike at 1800 nm in Figure 6.9b–d. Boxcar smoothing (Figure 6.9b) spreads the effect of the noise spike equally over the convolution interval. Polynomial smoothing (Figure 6.9c) spreads the effect over the region of fit but in the process produces features that fall below the normal spectrum. Both boxcar and polynomial smoothing had an effect only within the convolution and region of fit, respectively. Fourier smoothing (Figure 6.9d), on the other hand, produces smoothing with very little effect on the neighboring points. In this case, 39 coefficients produced the best smoothing without excessively altering neighboring points. It is only when the number of coefficients approaches 100 that the effect on neighboring points becomes a problem.

6.4.2 DECONVOLUTION VIA FOURIER SPACE

FSD is the process of removing the effect of the convolution function $g(k + \lambda)$ (Figure 6.6b) from the spectrum in Figure 6.6c. Another way of putting it is to say FSD allows you to take Figure 6.6c with its unresolved absorption bands and get Figure 6.6a, the original spectrum with its resolved bands. When you scan a spectrum with your machine, you may end up with a spectrum that looks very much like Figure 6.6c. You do not have spectrum (a) as a reference point. However, you can

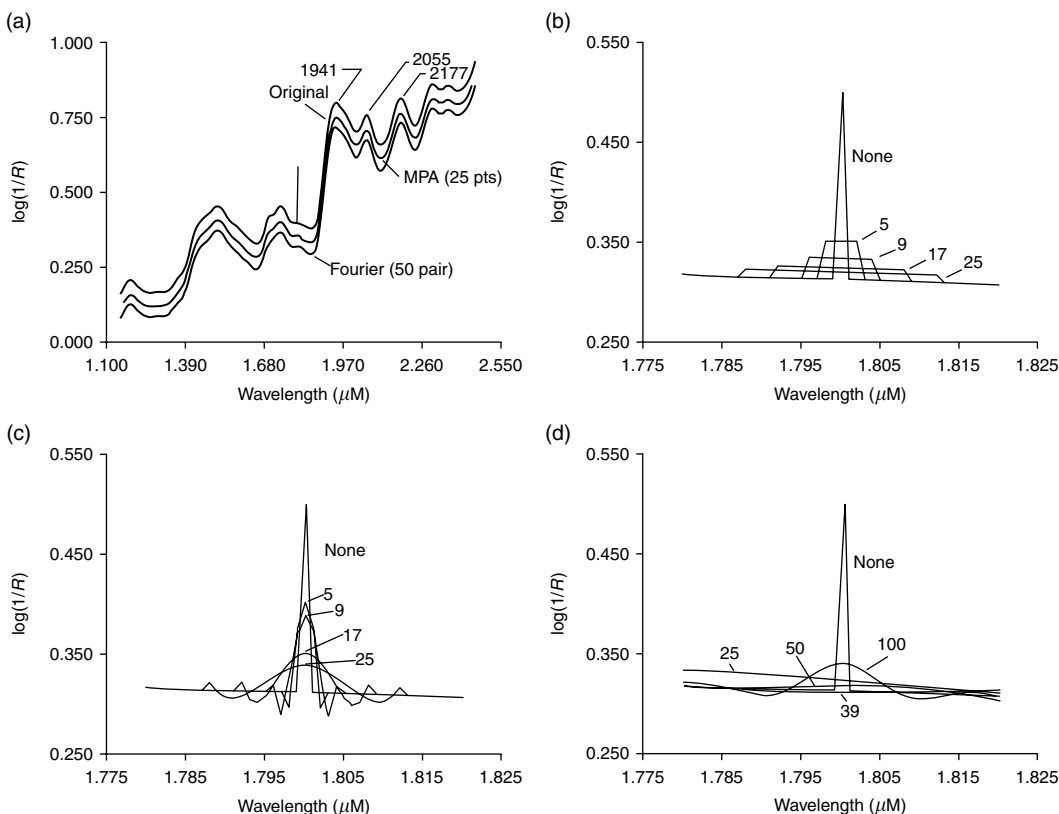


FIGURE 6.9 Spectra of wool: (a) Original spectrum with a noise spike at 1800 nm, MPA (25 pts) smoothed spectrum, and Fourier (50 pair) smoothed spectrum. (b) MPA smoothing of the noise spike. (c) Polynomial smoothing of the noise spike. (d) Fourier smoothing of the noise spike.

TABLE 6.3
Effect of MPA and Polynomial Smoothing on Position and Peak Value of Water Band in Wool

MPA smoothing			Polynomial smoothing ^a		
No. of points	λ_{max}	$\log(1/R)$	No. of points	λ_{max}	$\log(1/R)$
1	1941	0.7168	1	1941	0.7168
5	1941	0.7168	5	1941	0.7168
9	1942	0.7158	9	1941	0.7168
17	1942	0.7147	17	1941	0.7168
25	1943	0.7115	25	1941	0.7168

^a Fourier smoothing gave the same results but without loss of end points.

see a shoulder on spectrum (c) that alerts you to a “hidden” absorber. You would like to know the center wavelength of that absorber. FSD will allow you to make that determination.

The process is a rather simple and fun to do. First, you take the Fourier transform of the spectrum (c) in Figure 6.6. Then you multiply the resulting Fourier coefficients D , point by point, by an

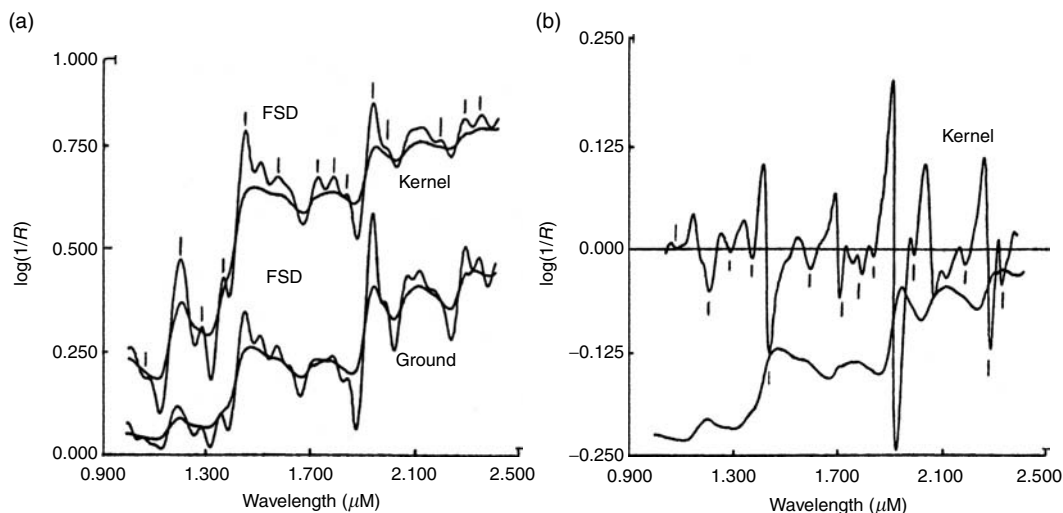


FIGURE 6.10 FSD (a) of the NIR spectra of whole-kernel and ground wheat along with the derivative spectrum of whole kernel wheat (b). Tick marks on the whole-kernel spectrum in (a) correspond to the tick marks on the negative peaks of the derivative spectrum in (b). Note the ease with which the deconvolved spectra in (a) compare with the original spectra.

exponential $e^{\beta z}$ (Figure 6.6e) where z is the Fourier index over the range from 0 to L . The value of β determines the degree of self-deconvolution. The Fourier transform of the result yields the self-deconvolved spectrum in (g) with its clearly resolved peaks.

It is even more fun to look at the effects achieved with FSD of real spectra as in Figure 6.10. Figure 6.10a gives the spectra of ground and whole-kernel wheat along with their FSD spectra. The FSD spectrum of kernel wheat has tick marks on absorption bands that correspond to the tick marks on the negative peaks of the second derivative spectrum of kernel wheat (Figure 6.10b). As can be seen, all the bands that appear in the FSD spectrum also appear in the derivative spectrum. The principal difference is that the derivative spectrum is in a form not easily related to the original $\log(1/R)$ spectrum.

How do we know if we have over-deconvolved? Well, we don't. Neither do we know if our derivative transform produces spurious bands. Both FSD and the derivatization of recorded data are pretty much an art, not a science. One just has to try it and develop a feel for its effect on the data. But here are some guidelines: (a) FSD will produce negative lobes when you over-deconvolve a spectrum. If these lobes go below zero, you have probably over-deconvolved. (b) Look at the narrowest bands. If lobe production around the narrow bands look as if they should not be there, then you probably have over-deconvolved. (c) Finally, you can compare deconvolved spectra with derivative spectra. Until you get a feel for FSD you may want to use this security blanket.

6.4.3 COMPUTING DERIVATIVES VIA FOURIER SPACE

NIR researchers are aware that certain advantages may be achieved with derivative spectra. Removal of the offset and the dominant linear term from the spectral data is observed in the computation of derivatives and moving from $\log(1/R)$ to derivatives was one of the first attempts to correct for particle-size variations.

There are basically three different approaches to take in computing derivatives from experimentally acquired discrete data (a) by convolution, (b) by polynomial regression, and (c) via the FD. The first method, namely MPA convolution, is derived from an expansion of the Taylor series. The

Taylor series expansion of a function $y = F(\lambda)$ at $(\lambda_i + \Delta\lambda)$ about λ_i is

$$F(\lambda_i + \Delta\lambda) = F_i + F'_i(\Delta\lambda) + \frac{F''_i(\Delta\lambda)^2}{2!} + \frac{F'''_i(\Delta\lambda)^3}{3!} + \dots \quad (6.14)$$

where F_i is the ordinate corresponding to λ_i and $(\lambda_i + \Delta\lambda)$ is the region of convergence. The function at $(\lambda_i - \Delta\lambda)$ is similarly given by

$$F(\lambda_i - \Delta\lambda) = F_i - F'_i(\Delta\lambda) + \frac{F''_i(\Delta\lambda)^2}{2!} - \frac{F'''_i(\Delta\lambda)^3}{3!} + \dots \quad (6.15)$$

The first four terms of Equation (6.14) and Equation (6.15) are added to get an expression for F''_i , yielding

$$F(\lambda_i + \Delta\lambda) + F(\lambda_i - \Delta\lambda) = 2F''_i + F_i(\Delta\lambda)^2 \quad (6.16)$$

Therefore

$$F''_i = \frac{F(\lambda_i + \Delta\lambda) - 2F_i + F(\lambda_i - \Delta\lambda)}{\Delta\lambda^2} \quad (6.17)$$

Equation (6.17) is called the first central difference [15] approximation of the second derivative of the function $F(\lambda)$ at λ_i , and the solution is simply an algebraic expression. Therefore, the second derivative of the absorbance $d^2A/d\lambda_i^2$ can be approximated by the following expression

$$\frac{d^2A}{d\lambda_i^2} = \frac{(A_{(\lambda_i+\Delta\lambda)}) - 2A_{\lambda_i} + (A_{(\lambda_i-\Delta\lambda)})}{\Delta\lambda^2} \quad (6.18)$$

where $\Delta\lambda$ is a small wavelength interval, say, 1 or 2 nm. Equation (6.18) is the mathematical representation of a simplified form $(A - 2B + C)$ extensively referred to in the literature [7]. As such, both are only an approximation of the true derivative.

Derivatives are inherently noisy. Choosing to do derivatives with only three points, as is implied by Equation (6.18), will result in noisy data. For example, if one were to compute the derivative of the noise spike in Figure 6.9, the results would be

$$\frac{d^2A}{d\lambda_i^2} = (V_{1799}) - 2(V_{1800}) + (V_{1801}) \quad (6.19)$$

or

$$(0.3134) - (2)(0.50) + (0.3130) = 1.6264 \quad (6.20)$$

a value that is not the derivative at 1800 nm, assuming the noise spike were not present. In practice, smoothing is incorporated into the derivative calculation by requiring values of A , B , and C in the simplified equation to be the computed averages of a segment of points [7].

Polynomial smoothing requires that an odd set of points be fitted to a polynomial and the center point calculated from the polynomial. Likewise, the derivative spectrum may also be computed from that same polynomial. For example, Equation (6.13) can be used to compute the second derivative of $F(\lambda)$ by noting that

$$F''(\lambda) = a_1 + 2a_2\lambda + \dots + na_n\lambda^{n-1} \quad (6.21)$$

The procedure is repeated for each group at a time, dropping one point on the left and picking up one at the right each time. As was the case with polynomial smoothing, derivative computations with polynomials can be enhanced by considering it a convolution process and using the convoluting numbers provided by Savitzky and Golay [6].

6.4.3.1 Fourier Derivatives

Computation of derivatives of NIR spectra in Fourier space is much more efficient than the computations in wavelength space. From Equation (6.3), letting $K = 2\pi k/N$, the first derivative with respect to n becomes

$$\frac{df(x)}{dx} = - \sum_{k=1}^{M-1} K a_k \sin(Kx) + \sum_{k=1}^{M-1} K b_k \cos(Kx) - \pi a_M \sin(\pi x) \quad (6.22)$$

The second derivative becomes the derivative of Equation (6.7)

$$\frac{df^2(x)}{dx^2} = - \sum_{k=1}^{M-1} K^2 a_k \cos\left(\frac{2\pi kx}{N}\right) - \sum_{k=1}^{M-1} K^2 b_k \sin\left(\frac{2\pi kx}{N}\right) - \pi^2 a_M \cos(\pi x) \quad (6.23)$$

Derivative spectra in wavelength space may be produced by applying the *inverse* Fourier transform (Equation (6.9) to Equation (6.22) and Equation (6.23)). Note that the position of first two terms in Equation (6.22) must be switched before doing the inverse transformation.

Figure 6.11a–c gives a comparison of the derivatives from three computational methods: (a) MPA a, (b) polynomial b, and (c) Fourier c. The MPA smoothing process used a segment size of 21 points, the polynomial segment size was 25, and the number of coefficient pairs used in the Fourier computation was 69. The Fourier derivatives contain the same number of points as the original spectrum while the convolution techniques drop points on both ends of the spectrum.

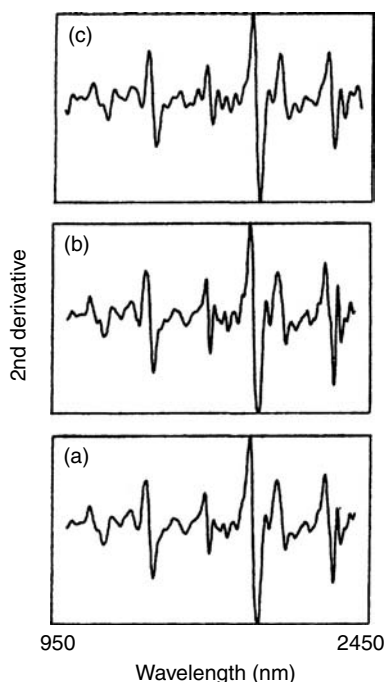


FIGURE 6.11 Comparison of derivatives obtained by MPA (a), polynomial (b), and Fourier (c) methods. There is very little difference except the Fourier methods give you all 1300 points while the other two methods drop points from each end of the spectra.

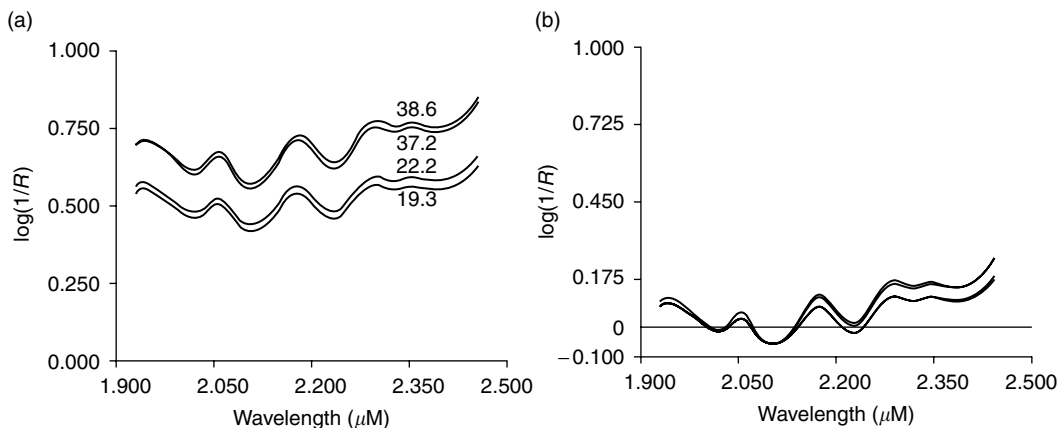


FIGURE 6.12 Relationship of the mean term a_0 to particle size: (a) Inverse transformed spectra using coefficients 1 to 99 and (b) inverse transformed spectra using coefficients 2 to 99 (leaving out the first term); note that the spectra come closer together in (b).

In view of the fact that the derivative spectra are made up of values computed from $K^z a_k$ and $K^z b_k$, it is interesting to contemplate the meaning of a derivative given a z value equal to 1.5 or some other fractional value.

Such derivatives are easy to compute in the FD, but their effects as well as their meaning are yet to be determined.

6.4.4 CORRECTS FOR PARTICLE SIZE ANOMALY

The first term a_0 in Equation (6.3) is the *mean* term. Exclusion of the mean term a_0 from the inverse Fourier transform makes a correction for linear background effects, the largest of which is particle size. This is illustrated in Figure 6.12 where the curves are offset due to particle size. Figure 6.12a shows four spectra of wool over the range from 1930 to 2450 nm, each sample having a different fiber diameter from 19.3 to 38.6 μm . Note that as the fiber diameter becomes smaller the spectra shift downward. Removing the mean term from the arrays of Fourier coefficients and applying the inverse Fourier transform gives the results in Figure 6.12b. The curves come closer together by virtue of excluding the mean term. A slight multiplicative difference still remains between the curves in Figure 6.12b. It is not possible to remove this difference entirely due to absorption distinctions within each sample.

Particle-size correction may not work to an advantage. Such corrections often produce poorer calibrations than without corrections. This is also true in wavelength space where derivative calibrations (which appear to correct for particle size) are found to give poorer results than $\log(1/R)$ calibrations. So questions arise concerning the advisability of particle-size corrections. Here again the Fourier approach has an advantage. With calibrations taking only 1.9 min (for 100 samples), the user is encouraged to test the calibration both ways, with and without the mean term.

6.5 QUANTITATIVE SPECTRAL ANALYSIS

6.5.1 ESTIMATING COMPOSITION WITH FOURIER COEFFICIENTS

Selecting the *optimum* wavelengths for calibration equations from 700 or so points in the spectra remains a problem. Spectral data are intercorrelated to a high degree, making it even more difficult

to mathematically pick out the *best set* wavelengths. This problem is referred to by statisticians as *multicollinearity* [8]. Even modern multivariate manipulations are not sufficient in and of themselves to yield the most robust calibration [9].

Combing, shown in Table 6.4, is a rather interesting way to demonstrate multicollinearity. Table 6.4 shows the effect of combing on the standard error of performance (SEP). Starting with 200 spectra with 840 points over the range 910 to 2588, we removed equally spaced segments of the data resulting in spectra with 99, 50, 25, 13, 7, and 5 points per spectrum in seven different datasets. The first spectral data point at 910 was retained in all sets, and the size of the comb was purely arbitrary. The results were surprising. Only the constituent nicotine showed a consistent degradation from an SEP of 0.178% for 840 points per spectrum to 0.292% for only five equally spaced points per spectrum. For all the datasets, there was sufficient information to give a good appraisal of the six constituents in tobacco. That is, performances of the calibrations with reduced data, even with only five points per spectrum, were not drastically different from the calibrations based on the original dataset with 840 points in each spectrum.

In Table 6.5, we see that the principal component transformation of the spectral data does not improve the performance of calibrations [10] for (DCM) extractables in wool. This table compares

TABLE 6.4
Effect of Combing on Performance (SEP) of NIR Calibrations for Six Constituents on Tobacco^a

No. of points	Standard errors of performance					
	NIC	SUG	NIT	WSN	POT	CAL
840	0.178	0.998	0.142	0.143	0.169	0.195
99	0.190	0.947	0.131	0.132	0.184	0.171
50	0.227	0.902	0.133	0.145	0.193	0.222
25	0.252	0.919	0.126	0.140	0.117	0.205
13	0.267	0.858	0.128	0.140	0.182	0.246
7	0.281	0.984	0.145	0.149	0.181	0.250
5	0.292	1.107	0.147	0.152	0.272	0.293

^a Number of samples for calibration = 100; number of samples for validation = 100.

TABLE 6.5
Comparison of PCR with MLR of $\log(1/R)$ Data

Model	R^2	SEC	SEP ^a
Forward MLR	.818	0.071	0.125
Backward MLR	.825	0.070	0.069
PCR			
1st 8	.714	0.088	0.253
All 19	.826	0.071	0.119
Branch-and-bounds			
Meth. = C	.823	0.070	0.063
Meth. = R	.826	0.071	0.064
Meth. = A.R.	.825	0.070	0.064

^a Residual grease in wool.

TABLE 6.6
Comparison of the Fourier Method with the Wavelength
Method for Estimating Sugars in Tobacco

Model	Terms	R^2	SEC	SEP
$\log(1/R)$	6	.968	0.690	0.998
DR/R	7	.972	0.644	0.990
$D2[\log(1/R)]$	7	.964	0.734	1.047
FF(1–11)	12	.964	0.838	0.975
FF(2–12)	12	.965	0.749	0.836

forward MLR, and backward MLR, with the branch-and-bounds and principal components regression (PCR). Interestingly, principal components was no better than backward MLR, not much better than forward MLR, and not nearly as good as the branch-and-bounds method of selecting wavelengths for the MLR algorithm. Although the forward linear regression performed well for calibration, the validation statistics (SEP) are inferior. Backward regression compares well with the branch-and-bounds method except for the additional wavelengths required to achieve comparable values of standard error calibration (SEC) and SEP. The branch-and-bounds program allows the choice of R^2 , adjusted R^2 , or Mallows's C_p statistics [11] as the criterion for selecting the best equation. The NIR team at the Wool Research Organization of New Zealand is convinced that PCR has no advantage and that the branch-and-bounds method is adequate for their work. These findings raise a real question: What is the best way to get a good, lasting calibration?

It was shown previously that the $\log(1/R)$ spectra of many agricultural products and food can be reproduced from the first 50 pair of Fourier coefficients including the mean term [4,5]. It seems logical to assume that if these coefficients can be used to reproduce the original spectra with such accuracy composition measurements could be made with those same coefficients. We will now demonstrate that calibration equations can be developed from the first six pairs of coefficients with no loss in accuracy.

Table 6.6 compares the SEC and SEP using Fourier coefficients with analyses using WD data. The dataset is based on 200 spectra of pulverized tobacco — 100 for calibration and 100 for testing the calibration. The average triplicate analyses of the sugars was attached to the spectral data for both calibration and prediction samples. The results show certain advantages in using Fourier coefficients for determining sugar levels in tobacco. First, the calibration process did not have to deal with the problem of selecting wavelengths; the first 12 or so coefficients were fitted into an MLR model similar to the procedure followed in the WD. Transformation to the FD arranges the *composition data* so that the most important information is in the first few coefficients. As long as the original spectra do not have very sharp peaks, as is the case for spectra of agricultural products, the first coefficients in the Fourier spectrum are the prime candidates for calibration equations. Second, the calibration procedure took far less time because the MLR algorithm did not have to search through 840 points.

We should not that the Fourier equation FF(1–11) included the mean term while the FF(2–12) equation excluded it, the latter having the lowest SEP (0.836%) for all three models. The equilibrium moisture content of tobacco is known to be proportional to sugar level. Moisture content is known to affect the particle-size distribution during grinding. Therefore, the improvement noted in Table 6.6 by leaving out the mean term is likely due to the particle-size correction achieved by excluding the mean term.

6.5.2 CUTTING COMPUTER STORAGE REQUIREMENTS BY 98%

Storing NIR spectral data can very quickly become a critical problem. For example, storing 200 spectra, each spectrum containing 1680 data points, on an IBM PS/2 operating under MS DOS

TABLE 6.7
Magnetic Storage Requirements for 200 Spectra
with Each Spectrum Containing 1680 Data Points

Data type	Pts/spectrum	Bytes stored	Space savings (%)
$\log(1/R)$	1680	1,359,616	0
F099	99	105,216	92
F012	12	28,416	98

will require more than 1.3 MB of magnetic storage (refer to Table 6.7). Adequate reproduction of agricultural spectra requires the retention of as few as 50 pairs of coefficients or 99 numbers, giving a savings in storage space of at least 92%. More than 2000 spectra can be archived as Fourier coefficients on a single 3.5 in. diskette. Since only 99 numbers per spectrum would have to be transmitted, modem transfer of spectra between locations is extremely fast.

If the data are to be used only for computations of chemical composition, magnetic storage requirements may be reduced even further. It has been demonstrated that fewer than six pairs of Fourier coefficients are needed to compute certain constituents in an agricultural product [4,5]. This results in a 98% savings in magnetic storage and means that the spectra from more than 9000 samples can be retained on a single 3.5 in. diskette.

The downward trend in the cost of magnetic storage media appears on the surface to diminish any concerns over data storage. Yet working with large datasets in the WD clearly demonstrates the need for an alternative method for archiving NIR data. Frustrations quickly emerge in the process of trying to determine the appropriate data type ($\log(1/R)$, derivatives, etc.) to be used for calibration equations. While removable hard disks and tapes provide attractive means of storage, the problems associated with cataloging the content of the disks becomes very frustrating. Compressing of data using Fourier transforms so that they will fit nicely on single diskettes is an attractive storage alternative.

6.5.3 REDUCING CALIBRATION TIMES BY 93%

Development of a ten-term calibration from 200 spectra with each spectrum containing 1680 points takes 10.5 min on an IBM PS/2 model 80 running at 25 mHz. In Fourier space a ten-term equation can be computed in 45 sec. This represents a 93% savings of computer time. Although these benchmarks were obtained on a system with a high-speed disk, it is readily apparent that the Fourier method would make 3.5 in. diskette systems tolerable even for large datasets. This brings up the next advantage.

6.5.4 ENCOURAGING CALIBRATION MAINTENANCE

Because it takes typically less than 45 sec to compute a calibration equation in Fourier space, NIR equipment can become an integral part of many more time-constrained quality control schemes. As new calibration samples become available the question as to when it would be proper to recalibrate is less likely to arise. Calibration is performed quickly and attention is directed to other things.

Table 6.8 illustrates the extent to which calibration maintenance can be taken. The table shows what the user must consider in the second year of a calibration used to estimate the neutral detergent fiber content of forages. Adjustments for bias are not taken into consideration here. The second line in the table shows that if the first year calibration is used to estimate the samples from the second year the standard error may jump to 3.912%. If he recalibrates (line 1 in Table 6.8), the SEP is 2.799%. Line 4 indicates that a 50/50 recalibration would be best. Line 5 indicates the hazard of using too few samples in the calibration; the R^2 of calibration is high but the SEP is poor. The computing time taken to run all eight calibrations was less than 6 min.

TABLE 6.8
Calibration Maintenance for Neutral Detergent Fiber in
Forges: Second Year Performance of NIR Calibration

Line	Composition of calib. file	Cal. <i>N</i>	<i>R</i> ²	PRED		
				SEC	<i>N</i>	SEP
1	100/2 ^a	134	.927	1.935	134	2.799
2	100/1	150	.930	1.806	134	3.912
3	75/1 + 24/2 ^b	147	.923	1.910	100	2.711
4	50/1 + 50/2	142	.912	2.120	100	2.410
5	25/2	34	.968	1.407	100	3.514
6	50/2	67	.938	1.952	134	2.391
7	100/1 + 25/2	182	.916	1.941	231	3.175
8	100/1 + 50/2	215	.899	2.183	199	2.609

^a Means 100% new samples from year 2 in the calibration file.

^b Means 75% samples from year 1, 25% from year 2.

6.5.5 REDUCING MULTICOLLINEARITY

Transformation of NIR spectra to the FD reduces multi-collinearity and the FD data still contain sufficient information to reproduce the spectra with very little error. Figure 6.13a–c illustrates the multicollinearity problem in the WD and compares it with multicollinearity in the FD. Note that the correlation coefficient for spectral data points separated by 2 nm across the wavelength range (based on 200 spectra) remains at 1.0 (Figure 6.13a across the entire range). Basing the correlation on the points separated by 10 nm (Figure 6.13b) we see that the correlation begins the drop only slightly. For data points separated by 20 nm the correlation drops to a minimum of .66 at the long-wavelength end of the spectra. However, for Fourier spectra (Figure 6.13d) the correlation between adjacent points never exceeds .66, and the first 12 coefficients never exceed an intercorrelation of .44. Figure 6.13d includes a pairing of the real (sine) and imaginary (cosine) coefficients. Even if we separate the real and imaginary coefficients (Figure 6.13e and Figure 6.13f) the correlation between any adjacent numbers never exceeds .66. Note also that the correlation of many adjacent pairs in Figure 6.13d, Figure 6.13e, and Figure 6.13f are close, if not equal, to zero. The low multicollinearity of data in the FD could be the basis for the excellent standard errors of prediction achieved from FD calibration equations with 11 terms.

6.6 ARTIFICIAL INTELLIGENCE

6.6.1 SPECTRAL SEARCHING, MATCHING, AND COMPONENT IDENTIFICATION

Modern NIR instrumentation with the support of high-speed personal computers has the potential for performing artificial intelligence (AI). Spectral searching and matching (SAM), determination of instrument noise, and detection of instrument anomalies are among the easily adaptable features that can be run transparent to the user. Incoming raw materials can be checked for quality or impurities by comparing spectra from the incoming material to a set of reference spectra. SAM programs would enable instruments to advise low-skilled personnel of the product type and of the most appropriate calibration equations available for the product. In addition, online monitors can be programmed to alert operators of changes taking place in the finished product.

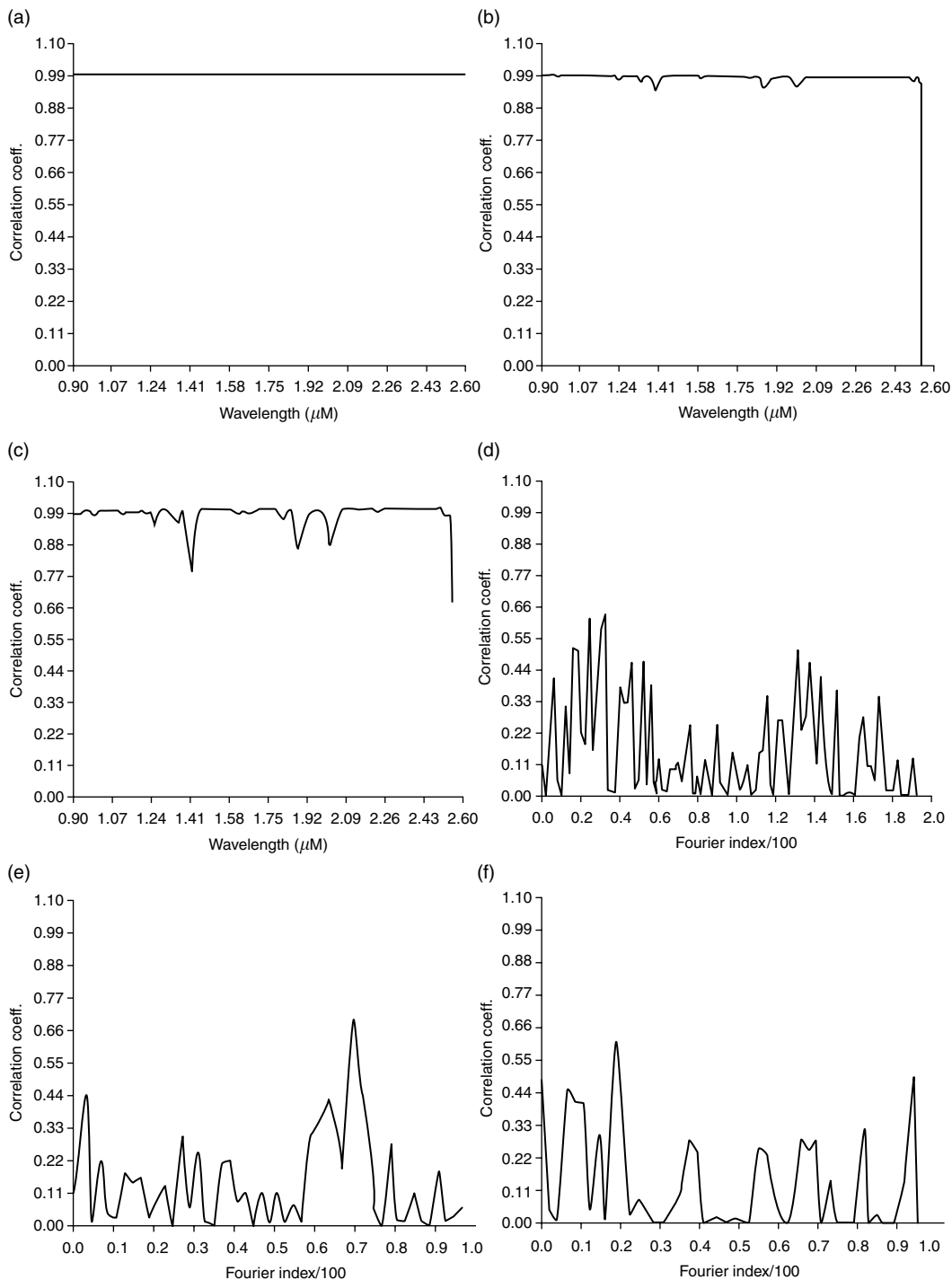


FIGURE 6.13 Comparison of multicollinearity in wavelength vs. Fourier space. Note that the intercorrelation of adjacent numbers in the Fourier spectra is much lower than that found in the WD.

TABLE 6.9
Results from Spectral Searching and Matching of Tobacco
Types with a Reference Base Consisting of Flue-Cured,
Burley, Maryland, and Dark-Fired Tobaccos

Type	FO	D1	D2	PW
FLU	100	100	100	99
DRK	100	100	100	100
MAR	93	92	92	38
BUR	94	94	97	84

Table 6.9 illustrates the power of searching in the FD. The reference database consisted of four spectra of flue-cured, burley, Maryland, and dark-fired tobaccos. Each spectrum was the average of 50 spectra. Then 50 additional samples, different from the ones that made up the average, from each type were compared against the database. The correlation coefficient (sometimes referred to as the cosine vector of the comparison) of each comparison was the basis of the match. The results in Table 6.9 give the percentage of the correct matches for the search. Fourier (FO) showed a superior performance over the search by first and second derivatives in the WD.

Work in mid-infrared and other analytical technologies [2,12,13] would lead us to believe that spectral searching and matching would be more efficiently implemented in the FD. Indeed this is true. In view of previous results [4,5] showing that as few as 50 pairs of Fourier coefficients can be used to approximate spectra with a high degree of accuracy, we have shown that spectral searching and matching in the FD can be achieved easier, faster, and without loss of precision and accuracy.

Furthermore, it is likely, based on recent work in the United Kingdom [14] and in this lab, that components in a mixture can be quickly identified by matching Fourier vectors rather than wavelength spectra. It has been demonstrated that as few as 25 pairs of Fourier are sufficient to make determinations of components in foods. Since only a few numbers need to be matched for each feed component, the procedure would be very timely.

6.6.2 CHECKING INSTRUMENT NOISE IN REAL TIME

Noise margins for certain NIR applications may need to be very low in order to get good performance. If these instruments are operating online, there needs to be a way to keep track of the noise levels. Analyses via the FD can be very useful in monitoring the performance of an instrument.

The noise spectrum in Figure 6.14 was obtained by reconstructing a spectrum from 100 pairs of Fourier coefficients and subtracting the reconstructed spectrum from the original. The standard deviation of the difference in this case is 82.8 μ OD. The standard deviation spectrum could serve as a real-time index of the noise. It could be recorded for every scan taken and a program could be written to alert the operator if the limits were exceeded.

6.6.3 TESTING FOR INSTRUMENT ANOMALIES IN REAL TIME

Mechanical problems in NIR instrumentation can be a real problem, especially for online instrumentation where environmental noise may defect our ability to hear audible noise that would normally be detected in a quiet laboratory. Mechanical noise, causing the lamp filament to vibrate, can be transferred to the spectrum. Information in the Fourier spectrum is a good indicator of this type of noise. Because the power spectrum is easily computed from the following:

$$PS = \sqrt{a_k^2 + b_k^2} \quad (6.24)$$

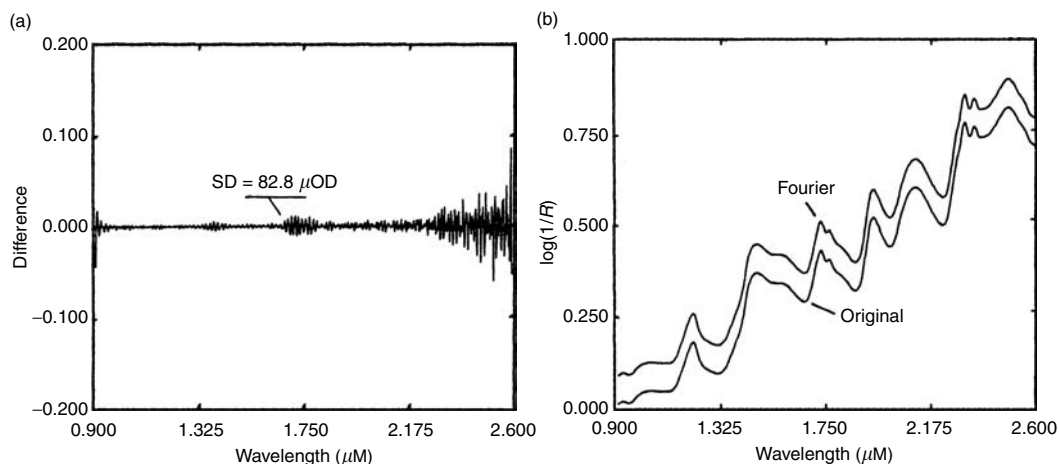


FIGURE 6.14 Routine noise information obtained using Fourier analysis. The Fourier derived spectrum was obtained by taking the inverse Fourier transform of the Fourier spectrum truncated to 100 pairs. Acquisition of this data could be transparent to the user of NIR instrumentation. The noise spectrum (b) was obtained by subtracting the original spectrum from the Fourier-reconstructed spectrum (a).

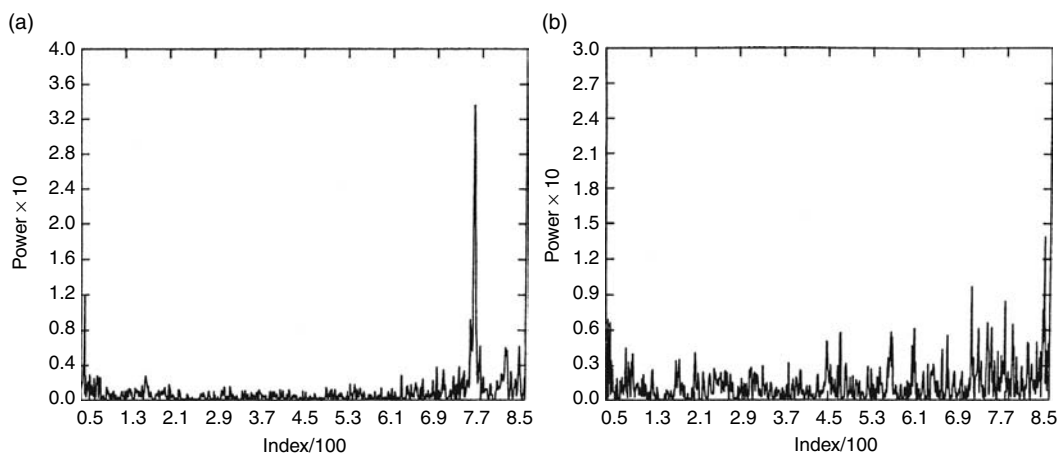


FIGURE 6.15 Power spectra of two scanning NIR spectrometers: (a) Comp/Spec No. 1 at North Carolina State University and (b) Comp/Spec No. 2 at the Oxford Experiment Station, Oxford, North Carolina.

a record of the performance of an instrument can be easily maintained. For agricultural spectra, the magnitude of PS above the Fourier index 200 is indicative of noise. If mechanical noise is present it will usually show up in this region. The increased power below index 200 accounts for the major part of the spectral information.

Figure 6.15a and Figure 6.15b gives the power spectrum of two Comp/Specs. Both systems were constructed identically. However, on examination of the power spectra of the two instruments a noise peak at index 757 was discovered for instrument 1, a peak that did not appear in the power spectrum of instrument 2. The peak at index 757 corresponds to a noise frequency of 4.8 Hz, a mechanical noise problem in one of the gears of the instrument.

Furthermore, it can be seen in Figure 6.15 that the magnitude of the noise below index 200 is much lower for instrument 1 than it is for instrument 2. Thus, one would consider instrument 1 to give better results than instrument 2 by virtue of its low noise.

6.7 INTERFEROGRAM SPACE

6.7.1 ESTIMATING CHEMISTRY FROM INTERFEROGRAM DATA

That the first 12 Fourier coefficients contain sufficient information to allow one to determine chemical constituents in many products has turned out to be a major advantage. Coupled with other research demonstrating that mid-IR interferograms could be used in searching and matching algorithms, it was our conviction that the chemical absorption information in the wavenumber domain spectrum of a sample obtained with an interferometer would be present in its interferogram. With this in mind, a low-resolution Fourier transform near-infrared (FT-NIR) spectrometer was designed and constructed using water-free quartz/inconel beam splitter, PbS detector, and a moving mirror on ball bearings that was motor driven.

Interferograms containing 512 points each were acquired in the diffuse reflectance mode from 66 samples of tobacco that had been analyzed for total alkaloids in advance. The samples were also scanned in a dispersion-type computerized spectrometer (Comp/Spec) to obtain $\log(1/R)$ spectra in the WD. The spectra from the 66 samples were split into two files, one containing the odd-numbered and the other containing the even-numbered spectra. Calibrations were developed on the odd spectra and validation was performed on the even spectra.

A typical interferogram of tobacco is shown in Figure 6.16a. All 512 points of the scan are shown. However, note that the interferogram numbers beyond 100 points contain very little information. The

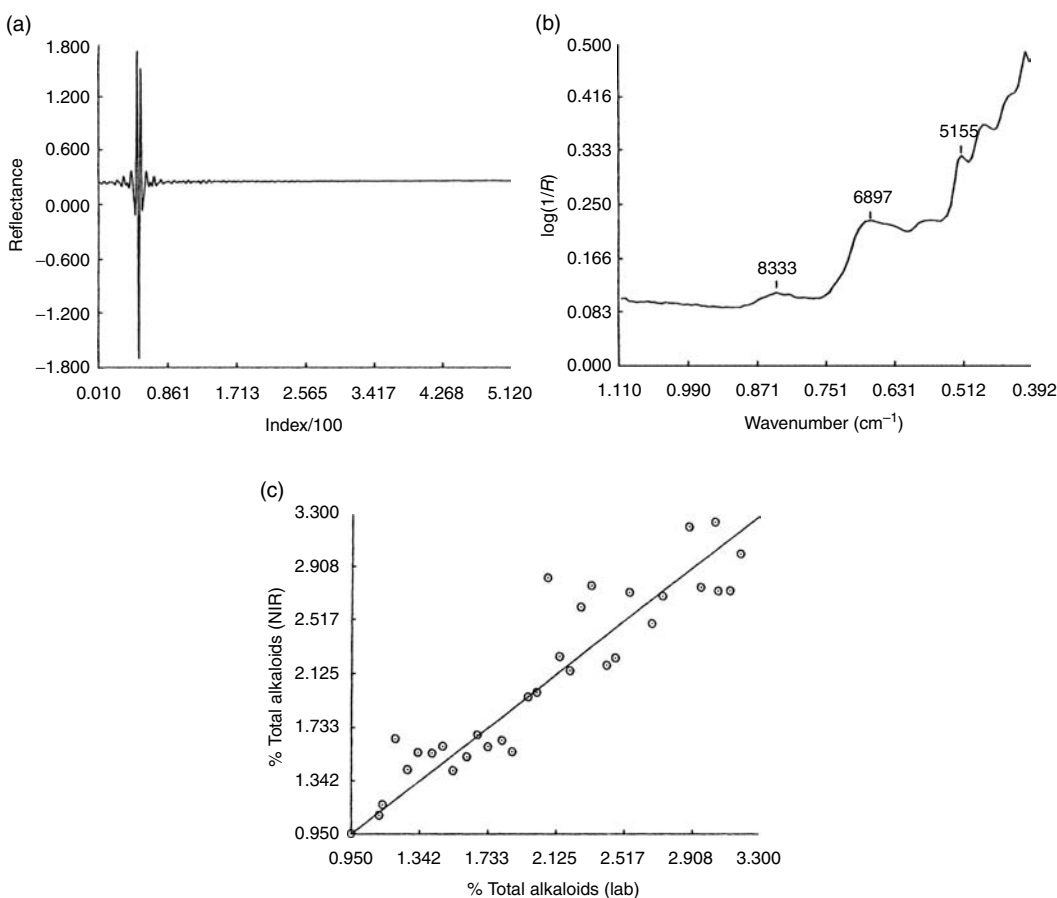


FIGURE 6.16 (a) Interferogram of tobacco. (b) $\log(1/R)$ spectrum produced from the interferograms shown in Figure 6.4 and Figure 6.5. (c) Predicted vs. actual total alkaloids in 33 samples of tobacco.

TABLE 6.10
Performance Data for Stepwise Multilinear Calibration for Total Alkaloids in Tobacco Using Interferograms Compared to Wavenumber and Wavelength Domain Data

	β^a	R^2	SEC	SEP
A. Interferogram data				
<i>Point number</i>				
52	-0.000133			
48	0.000966			
60	-0.004998			
44	0.004251	.920	0.393	0.617
B. Wavenumber data				
<i>Wavenumber</i>				
11,102	-0.00541			
6,773	-0.00859			
5,412	0.00694			
10,483	0.00777	.857	0.491	0.894
C. Wavelength data				
<i>Wavelength</i>				
2202	-22.174			
1476	421.317			
1670	199.074			
1508	-615.089	.928	0.369	0.414

^a The regression coefficients.

wavenumber spectrum obtained by Fourier transforming this interferogram to wavenumber space can be seen in Figure 6.16b. The absorption bands of water at 8333 cm^{-1} ($1.2\text{ }\mu\text{m}$), 6897 cm^{-1} ($1.45\text{ }\mu\text{m}$), and 5155 cm^{-1} ($1.94\text{ }\mu\text{m}$) are easily observed on the plot.

Performance of the FT-NIR instrument for determining total alkaloids in tobacco is given in Figure 6.16c. Four I-numbers within ten points of the center burst were selected by the multilinear calibration algorithm for determining total alkaloid composition (see Table 6.10A). These results support earlier findings that low-resolution information can be used to make chemical determinations in chemically complex samples. The SEC was 0.393% and the SEP was 0.617%. This is acceptable when compared to WD results (Table 6.10C, where the SEC was 0.369% and the SEP was 0.414%) obtained with the dispersion-based instrument. The wavenumber results (Table 6.10B) were slightly poorer (SEC = 0.491 and SEP = 0.894%). Keep in mind that the resolution of the FT-NIR instrument was greater than 60 cm^{-1} (25.3 nm at $2.0\text{ }\mu\text{m}$) while that of the Comp/Spec was better than 10 nm across the whole spectrum.

6.7.2 NO-MOVING-PARTS FT-NIR

Kawata and Minami [3] are doing some fascinating work with no-moving-parts FT-NIR. Their instrumentation is a combination of FT interferometry and multichannel (MC) detection and is called MCFT. It is well suited for measurements in the NIR, the range of which is limited only by the lack of suitable array detectors to cover the entire range. As it is, they can acquire spectra from 900 to $2000\text{ }\mu\text{m}$. Because their designs contain no moving parts, their instruments are compact, easy to align, very durable, and quite rugged.

The potential of interferograms in analytical chemistry [17] will increase in popularity in the years to come. No-moving-parts FT-NIR instruments will make this a much needed exploitation.

REFERENCES

1. Abdul Hamid, Personal communications.
2. D. A. Hanna, J. C. Marshall, and T. L. Isenhour, *Appl. Spectrosc.*, **32**: 314–317 (1976).
3. S. Kawata and S. Minami, Present status of spectroscopic instrumentation in Japan: Progress and development of MCFT-NIR, in *The Proceedings of the 2nd International NIRS Conference* (M. Iwamoto and S. Kawano, eds.), Korin Publishing Co. Ltd, Tsukuba, Japan, 1989, pp. 3–10.
4. F. G. Giesbrecht, W. F. McClure, and A. Hamid, *Appl. Spectrosc.*, **35**: 210–214 (1981).
5. W. F. McClure, A. Hamid, F. G. Giesbrecht, and W. W. Weeks, *Appl. Spectrosc.*, **38**: 322–328 (1984).
6. A. Savitzky and M. J. E. Golay, *Anal. Chem.*, **36**: 1627–1639 (1964).
7. P. C. Williams, Commercial near-infrared reflectance analyzers, in *Near-Infrared Technology in Agricultural and Food Industries* (P. Williams and K. Norris, eds.), American Association of Cereal Chemists, St. Paul, MN, 1987, p. 113.
8. E. O. Wesolowsky, *Multiple Regression and Analysis of Variance*, John Wiley & Sons, New York, 1976, pp. 49–64.
9. W. F. McClure, S. L. Ranford, and M. J. Hammersley, Analysis of near infrared data: Extracting information from a black hole, *Third International Conference on Near Infrared Spectroscopy*, Brussels, Belgium, June, 1990.
10. S. L. Ranford, M. J. Hammersley, and V. C. Patrick, Raw wool testing using near-infrared analysis, *Proceedings of the 7th International Wool Textile Research Conference: Vol. II*, Tokyo, 1985, pp. 167–176.
11. C. L. Mallows, *Technometrics*, **15**: 661–675 (1973).
12. H. B. Woodruff, G. L. Ritter, S. R. Lowry, and T. L. Isenhour, *Appl. Spectrosc.*, **30**: 213–216 (1976).
13. E. W. Small, G. T. Rasmussen, and T. L. Isenhour, *Appl. Spectrosc.*, **33**: 444–450 (1979).
14. A. M. C. Davies, CARNAC: A method of measuring composition by spectral searching and matching with Fourier vectors, Personal communications, 1990.
15. M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series 55, U.S. Government Printing Office, Washington, DC, 1964, p. 877.
16. W. R. Hruschka and K. H. Norris, *Appl. Spectrosc.*, **36**: 261–265 (1982).
17. R. M. Hoy and W. F. McClure, *Appl. Spectrosc.*, **43**: 1102–1104 (1989).

BIBLIOGRAPHY

SMOOTHING

- Aubanel, E. E. and K. B. Oldham. 1985. Fourier smoothing without the fast Fourier transform. *Byte Mag.*, (February): 211–218 [DOC 1385].
- Kauppinen, J. K., D. J. Moffatt, H. H. Mantsch, and D. G. Cameron. 1982. Smoothing of spectral data in the Fourier domain. *Appl. Optics*, **21**: 1866–1872 [DOC 1353].
- Nevius, T. A. and H. L. Pardue. 1984. Development and preliminary evaluation of modified Savitzky-Golay smoothing functions. *Anal. Chem.*, **56**: 2249–2251 [DOC 161].
- Porchet, P. and H. H. Gunthard. 1970. Optimum sampling and smoothing conditions for digitally recorded spectra. *J. Phys. E: Scientific Instr.*, **3**: 261–264 [DOC 362].
- Savitzky, A. and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, **36**: 1627–1639.
- Steinier, J., Y. Tremonia, and J. Deltour. 1972. Comments on smoothing and differentiation of data by simplified least squares. *Anal. Chem.*, **44**: 1906–1909 [DOC 54].
- Wilson, P. D. and T. H. Edwards. 1976. Sampling and smoothing of spectra. *Appl. Spectrosc. Rev.*, **12**: 1–81 [DOC 471].

FOURIER TRANSFORMS

- Bell, R. J. 1972. *Introductory Fourier Transforms*. Academic Press, New York.
- Brigham, E. O. 1974. *The Fast Fourier Transform*. Prentice-Hall, Englewood Cliffs, NJ.
- Champeney, D. C. 1973. *Fourier Transforms and Their Physical Applications*. Academic Press, New York.
- Jenkins, E. M. and D. G. Watts. 1969. *Spectral Analysis and Its Application*. Holden-Day, San Francisco.
- Marshall, A. G. (ed.). 1982. *Fourier, Hadamard, and Hilbert Transforms in Chemistry*. Plenum Press, New York.
- Rayner, J. N. 1971. *An Introduction to Spectral Analysis*. Pion LTD, London.
- Singleton, R. C. 1969. *IEEE Trans. Audio Electroacoust.*, AU-17(2): 93–103.
- Swartz, M. and L. Shaw. 1975. *Signal Processing: Discrete Spectral Analysis, Detection and Estimation*. McGraw-Hill, New York.

FOURIER TRANSFORM INFRARED SPECTROSCOPY

- Ferraro, J. R. and L. J. Basile. 1979. *Fourier Transform Infrared Spectroscopy*. Academic Press, New York.
- Griffith, P. R. 1975. *Chemical Infrared Fourier Transform Spectroscopy*. John Wiley & Sons, New York, p. 506.
- Hanna, D. A., J. C. Marshall, and T. L. Isenhour. 1976. *Appl. Spectrosc.*, 32: 314–317.
- Small, E. W., G. T. Rasmussen, and T. L. Isenhour. 1979. An Infrared search system based on direct comparison of interferograms. *Appl. Spectrosc.*, 33: 444–450.
- Woodruff, H. B., G. L. Ritter, S. R. Lowry, and T. L. Isenhour. 1976. Pattern recognition methods for the classification of binary infrared spectral data. *Appl. Spectrosc.*, 30: 213–216.

FOURIER ANALYSIS OF NEAR-INFRARED SPECTRA

- Davies, A. M. C. and W. F. McClure. 1987. CARNAC: Making chemical determinations by pattern recognition. *Proceedings of the 5th Pacific Scientific Users Conference*. Rockville, Maryland (May 4–6).
- Davies, A. M. C., S. M. Ring, J. Franklin, A. Grant, and W. F. McClure. 1985. Prospects of process control using Fourier transformed infrared data. *Proceedings of the 1985 International Conference on Fourier and Computerized Infrared Spectroscopy*. Ottawa (June 24–28).
- Giesbrecht, F. G., W. F. McClure, and A. Hamid. 1981. The use of trigonometric polynomials to approximate visible and near infrared spectra of agricultural products. *Appl. Spectrosc.*, 35: 210–214.
- McClure, W. F. and A. M. C. Davies. 1985. Fourier analysis of near-infrared data. *Proceedings of the 1985 International Conference on Fourier and Computerized Infrared Spectroscopy*. Ottawa (June 24–28).
- McClure, W. F. and A. M. C. Davies. 1987. Fourier self-deconvolution of near-infrared spectra of chemically complex samples. *Proceedings of the 6th FTS Conference*. Vienna (August 24–28).
- McClure, W. F., A. Hamid, F. G. Giesbrecht, and W. W. Weeks. 1984. Fourier analysis enhances NIR diffuse reflectance spectroscopy. *Appl. Spectrosc.*, 38: 322–328.
- Wesolowsky, E. O. 1976. *Multiple Regression and Analysis of Variance*. John Wiley & Sons, New York, pp. 49–64.

CONVENTIONAL NIR ANALYSIS

- Ben-Gera, I. and K. H. Norris. 1968. Direct spectrophotometric determination of fat and moisture in meat products. *J. Food Sci.*, 33: 64–67.
- Bittner, D. R. and K. H. Norris. 1968. Optical properties of selected fruits vs. maturity. *Trans. Am. Soc. Agric. Eng.*, 11: 534–537.
- Davies, A. M. C. and W. F. McClure. 1985. Near infrared analysis of foods. *Anal. Proc. (England)*, 22: 321–322.
- Devaux, M. F., D. Bertrand, and G. Martin. 1986. Discrimination of bread-baking quality of wheats according to their variety by near-infrared reflectance spectroscopy. *Cereal Chem.*, 63: 151–154.
- Gaines, T. P. and G. A. Mitchell. 1979. Chemical methods for soil and plant analysis. Univ. Georgia, Coastal Plain Exp. Sta. *Agronomy Handbook*, No. 1, pp. 65–70.
- Hamid, A. and M. L. McLester. 1977. Personal communications.
- Hamid, A., W. F. McClure, and W. W. Weeks. 1978. Rapid spectrophotometric analysis of the chemical composition of tobacco. Part 2: Total alkaloids. *Beitr. Tabakforsch.*, 9: 267–274.

- Hamid, A., W. F. McClure, and T. B. Whitaker. 1981. NIR analysis of food products: Part II. Stepwise linear regression in memory limited environment. *Am. Lab.*, 13: 108–121.
- Hruschka, W. F. and K. H. Norris. 1982. Least-squares curve fitting of near infrared spectra predicts protein and moisture content of ground wheat. *Appl. Spectrosc.*, 36: 261–265.
- McClure, W. F. 1968. Spectral characteristics of tobacco in the near infrared region from 0.6 to 2.6 microns. *Tobacco Sci.*, 12: 232–235.
- McClure, W. F. 1984. History and future prospects for near-infrared analysis. *Anal. Proc. (England)*, 21: 485–486.
- McClure, W. F. and R. E. Williamson. 1982. Rapid spectrophotometric analysis of the chemical composition of tobacco. Part 3: Polyphenols. *Beitr. Tabakforsch.*, 11: 219–227.
- McClure, W. F. and A. Hamid. 1980. Rapid NIR measurement of the chemical composition of foods and food products. Part 1: Hardware, *Am. Lab.*, 12: 57–69.
- McClure, W. F., K. H. Norris, and W. W. Weeks. 1977. Rapid spectrophotometric analysis of the chemical composition of tobacco. Part 1: Total reducing sugars. *Beitr. Tabakforsch.*, 9: 13–17.
- Norris, K. H. and W. L. Butler. 1961. Techniques for obtaining absorption spectra on contact biological samples. *IRE Trans. Biomed. Electron.*, 8: 153.
- Norris, K. H. 1964. *Trans. Am. Soc. Agric. Eng.*, 7: 240.
- Snook, M. E. and O. T. Chortyk. 1982. An improved extraction-HPLC method for tobacco polyphenols. *Tobacco Sci.*, 26: 25–29.
- Williamson, R. E., J. F. Chaplin, and W. F. McClure. 1986. Near-infrared spectrophotometry of tobacco leaf for estimating tar yield of smoke. *Proceedings of 40th Tobacco Chem. Res. Conf.*, Knoxville, Tenn.
- Williamson, R. E. and W. F. McClure. 1983. The role of near-infrared spectrophotometry in tobacco genetic research. *Abstracts, 10th Tobacco Annual Meeting Fed. Anal. Chem. and Spec. Soc.*, Philadelphia, p. 79.
- Williamson, R. E. and W. F. McClure. 1985. Near-infrared spectrophotometry in tobacco genetic research. *Proceedings of 39th Tobacco Chem. Res. Conf.*, Montreal, p. 7.
- Williamson, R. E. and W. F. McClure. 1986. Rapid spectrophotometric analysis of the chemical composition of tobacco. Part 4: Total nitrogen. *Tobacco Sci.*
- Williamson, R. E., V. A. Sisson, and W. F. McClure. 1986. Estimation of total nitrogen in tobacco tissue by near-infrared reflectance spectrophotometry. *Suppl. to Plant Phys.*, 80: 14.

7 NIR Spectroscopy Calibration Basics

Jerome J. Workman, Jr.

CONTENTS

7.1	Introduction	123
7.1.1	Basic Discussion.....	123
7.1.2	Instrumentation: NIR Sensitivity and Detection Limits	126
7.2	The NIR Method: Review of the Concepts	126
7.2.1	Relating Absorbance to Concentration	126
7.2.2	Major Error Sources in NIRS	128
7.2.2.1	Population Error	130
7.2.2.2	Laboratory Error	130
7.2.2.3	Packing Error.....	130
7.2.3	The Meaning of Outliers	130
7.3	The Calibration Process	132
7.3.1	Selecting the Calibration Set.....	132
7.3.2	Determination of Standard Concentrations: Manual Methods	134
7.3.3	Collecting Optical Data	134
7.3.4	Developing the Mathematical Model.....	135
7.3.5	Validating the Mathematical Model	137
7.3.6	Routine Analysis and Modeling.....	139
7.3.7	Transfer of Calibrations	140
7.3.8	Performing an Actual Feasibility Study	140
7.4	Statistical Terms	140
7.5	Test Statistics.....	141
7.5.1	Principal Components Regression and Partial Least Squares Testing	147
7.5.2	ASTM Practice for Multivariate Regression	148
	Acknowledgment	149
	References	149

7.1 INTRODUCTION

7.1.1 BASIC DISCUSSION

The purpose of this chapter is to provide a quick reference/overview to near-infrared (NIR) spectroscopy “calibrationists” on the important aspects of quantitative or qualitative calibration techniques as illustrated by the flow chart in Figure 7.1. The main task of the computer in NIR spectroscopy, aside from driving the instrument or collecting data, is to interpret the spectra using a variety of multivariate mathematical techniques. These techniques are used to produce a mathematical calibration model. Figure 7.2 illustrates the various sample sets used for testing calibration models and the conditions where recalibration is indicated. The purpose of the calibration model is

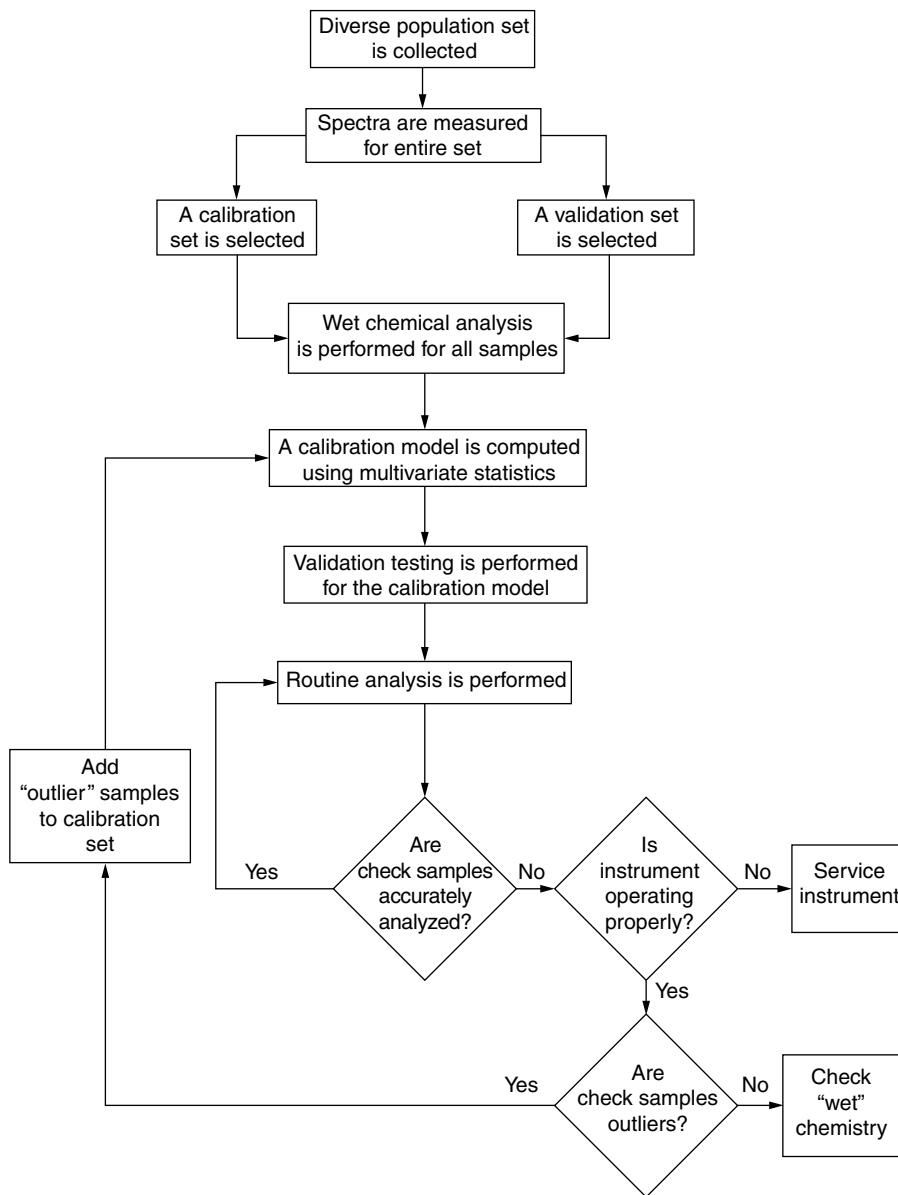


FIGURE 7.1 NIRS calibration flow chart.

to relate the concentration of some analyte found in a sample to the spectral data collected from that sample.

The task of calibration can be somewhat elementary when dealing with one- or two-component matrices but becomes quite complex when dealing with samples of biological origin where only minute changes in absorbance indicate large concentration changes, and where interfering bands overlap severely causing interference/interaction with bands representing the component(s) of interest. In biological samples of this nature, the detection limit of even major components can be as high as 0.5 to 1.0% absolute. In contrast, the detection limit for water in acetonitrile might be 50 or 100 $\mu\text{g/l}$ (0.005 to 0.01%). A great deal of NIR mythology surrounds such differences in performance in the use of the technique depending on the application. For example, the user analyzing a diverse product such as meat and bone meal may obtain a standard error for calibration of 1.1% absolute for protein

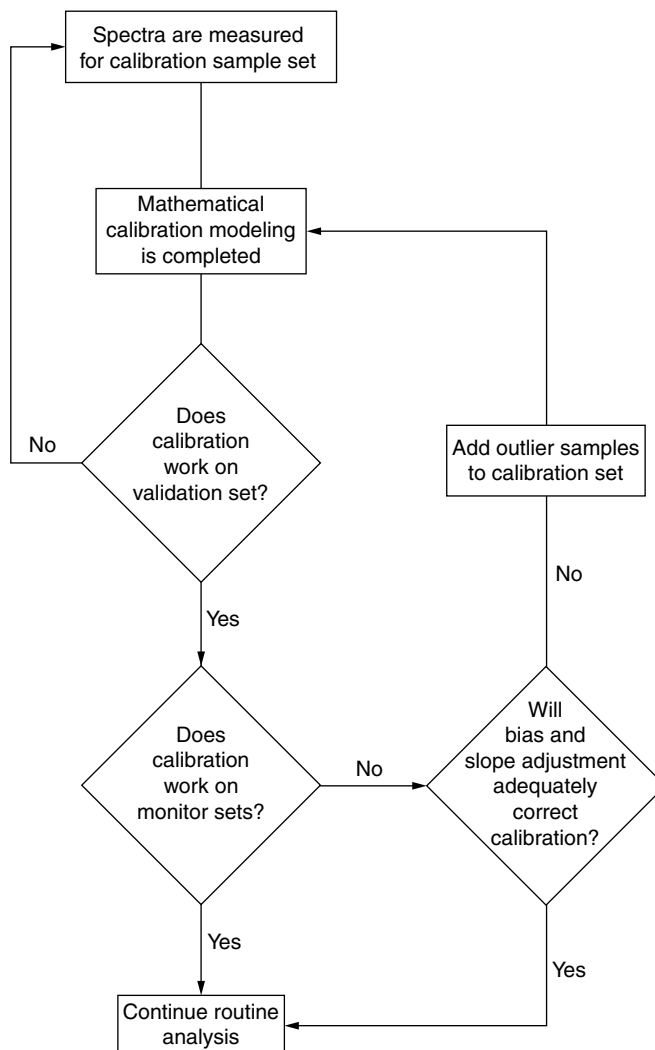


FIGURE 7.2 NIRS calibration validation and monitoring chart.

($N * 6.25$) using modified partial least squares with sophisticated sample selection, scatter correction, mathematical modeling, instrument standardization, and the like, whereas his or her counterpart in the chemical division of the company achieves a standard error for calibration of 0.05% for moisture in toluene using a single-term second derivative data pretreatment, a simple mixture design for calibration samples, and a two-wavelength stepwise regression. Now unless the two analysts are thoroughly trained in NIR technology they will argue about which method of calibration is best when they are simply involved in different uses of the analytical method. These are the types of problems that will be addressed briefly in this chapter.

In the present NIR spectroscopy technique, radiant energy in the 700 to 2600 nm portion of the spectrum is diffusely reflected or transmitted from or by a sample and detected using either a silicon photodiode detector (below 1100 nm) or a photoresist cooled or uncooled lead sulfide detector (above 850 nm). When light interacts with the sample, the amount of reflected or transmitted energy received at the detector is dependent on both the chemical (molecular absorbance) and physical (scattering/reflective) properties of the sample as well as the optical measurement geometry. These principles are best understood starting with the Fresnel, Raleigh, and Mie equations, and continuing

with discussions by Schuster, Kubelka, and Munk. An excellent discussion of these principals can be found in Reference 1.

When particle size becomes small and scattering large, reflectance losses will restrict the use of transmission measurements and decrease signal-to-noise in diffuse reflectance measurements. When path length is reduced to allow transmission, there is still the problem with sampling error in nonuniform samples. Reflectance losses still create a significant decrease in signal that is not related directly to chemical information. In cases where small particles are involved, and where surfaces are Lambertian (diffuse reflectors), reflectance measurements provide the greatest bulk chemical information. Diffuse reflectance measurement geometries tend to exclude a large portion of the specular component and thus are not affected to a large extent by surface reflectance/scattering losses. As samples increase in absolute reflectivity, less bulk information can be measured using either transmission or diffuse reflectance measurement geometries. As a sample becomes more like a mirrored surface (highly isotropic), only specular reflectance using s- or p-polarized light can yield information on surface properties.

NIR spectroscopy technology involves the multidisciplinary approaches of the analytical chemist, statistician, and computer programmer. Knowledge of reflectance and general optical spectroscopy is also essential in the advanced use of this technology. *Chemometrics* is a term applied to the generic discipline involving computers and mathematics to derive meaningful chemical information from samples of varying complexity. It is important for users managing NIR equipment to at least familiarize themselves with basic concepts from these disciplines.

7.1.2 INSTRUMENTATION: NIR SENSITIVITY AND DETECTION LIMITS

When defining instrument performance for a specific NIR spectroscopic application, the terms *detection limit* and *sensitivity* (or *signal-to-noise*) can be used. Detection limit can be loosely approximated for any NIR spectroscopic method as equal to three times the Standard error of calibration (SEC) for the specified application. The detection limit for a specific application may also be described as three times the repack error for the sample spectrum representing the lowest constituent concentration. This value approximates a 99% confidence that a difference in concentration between two samples can be correctly determined.

Sensitivity for any NIR method can be evaluated as the slope of the calibration line for concentration of analyte (y-axis), vs. change in optical response (x-axis), between samples of varying concentration. Sensitivity, from a purely instrumental point of view, is expressed as signal-to-noise or the ratio of peak height for a particular compound vs. peak-to-peak noise at some absorbance (usually zero).

7.2 THE NIR METHOD: REVIEW OF THE CONCEPTS

7.2.1 RELATING ABSORBANCE TO CONCENTRATION

The NIR spectroscopy algorithms used to “interpret” optical data for absorbing samples may be explained as different approaches to relating sample absorbance (A) at specific wavelengths to analyte concentrations via Beer’s law. To continue:

$$A = Mcd \quad (7.1)$$

where A = absorbance (optical density)

M = molar absorptivity

c = molar concentration of absorber

d = sample path length

and thus

$$c = \frac{A}{Md} \quad (7.2)$$

So the multiregression equation commonly used for calibration is

$$Y = B_o + B_i(-\log R_i)_N + E \quad (7.3)$$

where Y = percent concentration of absorber

B_o = intercept from regression

B_i = regression coefficient

i = index of the wavelength used and its corresponding reflectance (R_i)

N = total number of wavelengths used in regression

E = random error

This is actually a form of Beer's law with each B term containing both path length and molar absorptivity (extinction coefficient) terms. Most simply, the concentration is related to the optical data as

$$\text{Conc.} = \frac{\text{Change in concentration}}{\text{Change in absorbance}} \times \text{Absorbance} + \text{Some error}$$

or

$$\text{Conc.} = K \times \text{Absorbance} + \text{Some error}$$

Thus, K the regression coefficient is equal to the change in concentration divided by the change in absorbance. Therefore if there is a large change in concentration for the calibration set with a relatively large change in absorbance, the regression coefficients tend to stay small, indicating a large sensitivity and signal-to-noise ratio. In contrast, if we have a large concentration change relative to a small change in absorbance, the regression coefficients tend to be large and indicate low sensitivity and low signal-to-noise. This simple model applies whether a univariate (one-wavelength) or multivariate (many-wavelength) regression is performed.

If calibrations are performed for the simplest case involving one analyte per wavelength and no instrument error (i.e., no instrument drift with time, or noise), Equation (7.3) could be used with a single term.

$$Y = B1(-\log R) \quad (7.4)$$

By knowing Y , and measuring $-\log R$ at each wavelength i , Equation (7.4) could be used to solve for B . However, real-world complications exist, such as instrument noise, drift, nonlinearity between optical data and analyte concentration (deviations from Beer's law), scattering, nonlinear dispersion, error in reference laboratory results, physical property variations in samples, chemically unstable samples, band overlap, band broadening, and sampling errors. If ideal conditions are approximated (a) noise is assumed to be stochastic (random) and not unidirectional; (b) precautions are taken to optimize signal-to-noise ratios; (c) reference readings at each wavelength and collected to negate drift considerations; and (d) excellent laboratory technique and sample selection protocol are followed, Equation (7.3) may be used to solve for $B(o)$ and $B(i)$. This multilinear regression is a mathematical model relating the absorbance of several samples to their analyte concentration (as previously determined via primary wet chemical methods). The regression provides a best-fit linear model for absorbance vs. analyte concentration, and the mathematical model minimizes the sum of

the square residuals (distances) from each data point to the regression line (termed Y estimate). Note that Beer's law is a rigorously derived model applying only to transmission spectroscopy.

Beer's law applies where refractive index, scattering specular reflection at "infinite-finite" numbers of surfaces all obey the Fresnel formulas. A definite reflectance theory does not exist, as the convolution of an infinite number of integrals would be required to describe all the combined light interaction effects at all surfaces under varying conditions. Thus, Beer's law is often shown to illustrate the properties of NIR spectroscopy for lack of an ideal model.

So how does one generate a calibration? Let us go back to basics and observe that (ideally) one can find a suitable wavelength for any given substance wherein absorption is proportional to concentration (Beer's law). But as we noted, the real world is often nonideal, and absorbances deviate from Beer's law at higher concentrations due most often to nonlinearity of detection systems, scattering effects (which are wavelength-dependent), and stray light-caused nonlinearity. Note that spectra of compounds with absorbance bands more narrow than the instrument resolution can demonstrate substantial nonlinearities due to the stray light characteristics of the instrument.

Even when one restricts measurements to lower concentrations where the relationship between change in absorbance and concentration is linear, the calibration line seldom passes through the origin. Accordingly, even for single components, an offset is most often observed. The use of derivative math pretreatments is often used to reduce the nonzero bias, yet derivatives do not remove multiplicative error and nonlinearities. The offset occurs as a compensation to background interference and scattering. In a practical sense, most of the previous considerations do not matter except in an academic sense. This is due to the fact that the practical tools used in NIR are mostly based on empirical calibration methods where calibration techniques are well understood. Calibration equations using multivariate techniques are used to compensate for the full variety of common variations found in "noisy" chemical values and imperfect instrumental measurements. This is why properly formulated calibration models work extremely well despite the imperfect world of the analyst.

If a set of samples is analyzed with high precision by some reference (standard) method so the analyte concentrations are known, they can be used as a "teaching set" to generate an equation suitable for subsequent predictions. To be a good teaching set, the samples must evenly span the concentration range of interest. There is analytical "danger" in developing calibrations using sample sets with uneven constituent distributions as the mathematical calibration model will most closely fit the majority of samples in the calibration set. Therefore, a calibration model will be weighted to most closely fit samples at, or approximately at, the mean concentration value. Conversely, an evenly distributed calibration set will equally weight the calibration model across the entire concentration range. A properly developed calibration model will perform most accurately for samples at high and low concentration ranges when the calibration set is evenly distributed.

Note that the points do not lie in a straight line but are removed from a line by some distance (called a *residual*). With a mathematical treatment known as a *linear regression*, one can find the "best" straight line through these real-world points by minimizing the residuals. This line is known as the *calibration line*, and its equation can be used to determine the concentration of unknown samples. If we use principal components regression (PCR) or partial least-squares (PLS), we are regressing the sample scores rather than the optical data directly. A review of calibration problems encountered using standard spectroscopic techniques is outlined in the following paragraphs.

7.2.2 MAJOR ERROR SOURCES IN NIRS

In reflection of radiation at solid matte surfaces, diffuse and specularly reflected energies are superimposed. The intensity of the diffusely reflected energy is dependent on the angles of incidence and observation, but also on the sample packing density, sample crystalline structure, refractive index, particle size distribution, and absorptive qualities. Thus, an ideal diffusely reflecting surface can only be approximated in practice, even with the finest possible grinding of the samples. There are always

coherently reflecting surface regions acting as elementary mirrors whose reflection obeys the Fresnel formulas. Radiation returning back to the surface of the sample from its interior can be assumed as largely isotropic and should thus fulfill the requirements of the Lambert law. The assumption is made that radiant energy is continuously removed from the incident NIR beam and converted to thermal vibrational energy of atoms and molecules. The decrease in the intensity of the diffusely reflected light is dependent on the absorption coefficient of the sample. The absorption coefficient (K), when taken as the ratio K/S , where S is the scattering coefficient, is proportional to the quantity of absorbing material in the sample.

Utilizing the Kubelka–Munk (K–M) theory we then can relate the reflectance (R) to the absorption (K) and the scattering coefficient (S) by the equation:

$$\frac{K}{S} = \frac{(1 - R)^2}{2R} = F(R)$$

It may be stated that R , the diffuse reflectance, is a function of the ratio K/S is proportional to the addition of the absorbing species to the reflecting sample medium. On these relationships is based the assumption that the diffuse reflectance of an incident beam of radiation is directly proportional to the quantity of absorbing species interacting with the incident beam, and so R depends on analyte concentration.

NIR spectroscopic theory does not have to assume a linear relationship between the optical data and constituent concentration, as data transformations or pretreatments are used to linearize the reflectance data. The most used linear transforms include $\log(1/R)$ and K–M as math pretreatments. Calibration equations can be developed that compensate to some extent for the nonlinear relationship between analyte concentrations and $\log(1/R)$ or K–M-transformed data. PCR, PLS, and multilinear regression can be used to compensate for the nonlinearity.

If a matrix absorbs at different wavelengths than the analyte, K–M can prove to be a useful linearization method for optical data [2]. If the matrix absorbs at the same wavelength as the analyte, $\log(1/R)$ will prove to be most useful to relate reflectance to concentration. Attempts to minimize the effects of scattering and multicollinearity using standard normal variate and polynomial baseline correction are described in Reference 3. When generating calibration equations using samples of known composition, the independent variable is represented by the optical readings ($-\log R$) at specific wavelengths, while the analyte concentration (as determined by manual laboratory technique) is the dependent variable. The stepwise multiple regression statistic allows for the selection of calibration spectral features that correlate (as a group) most closely to analyte concentration for a particular sample set. Once optimal wavelengths are selected, the NIR spectroscopy instrument can be calibrated to predict unknown samples for the quantity of the desired analyte. Thus, regression analysis is used to develop the relationship (regression calibration equation) between several spectral features and the chemical analyte (constituent) being investigated. Note that calibration equations will also contain wavelength terms to compensate for repack variations and interferences such as sample moisture content.

Questions often arise as to which mathematical treatments and instrument types perform optimally for a specific set of data. This is best addressed by saying that reasonable instrument and equation selection composes only a small quantity of the variance or error attributable to the NIR analytical technique for any application. Actually, the greatest error sources in any calibration are generally reference laboratory error (stochastic error source), repack error (nonhomogeneity of sample — stochastic error source), and nonrepresentative sampling in the learning set or calibration set population (undefined error).

$$\text{Total variance in analysis} = \text{Sum of all variances due to all error sources}$$

Recent NIR research of necessity is moving toward an explanation of the chemistry involved in the various applications. Originally, the technology was totally empirical and only statistical methods were used to provide tests of calibration validity. With an increasing knowledge base, the NIR community is designing experiments in an attempt to reveal more information about this technology for samples with vastly differing chemical and physical properties. Thus, future users will be able to solve NIR spectroscopy problems with greater chemical and statistical knowledge a priori.

The largest contributions to calibration error can be minimized by reducing the major contributors to variance. These major error sources are described in Table 7.1 and in the following text.

7.2.2.1 Population Error

Population sampling error can be minimized by collecting extremely comprehensive datasets and then reducing them via subset selection algorithms (e.g., Bran and Luebbe PICKS Program, and NIRSystems Subset algorithm). These techniques allow maximum variation in a calibration set with minimum laboratory effort [4,5].

7.2.2.2 Laboratory Error

This source of error can be substantially reduced by performing an in-house audit for procedures, equipment, and personnel, paying particular attention to sample presentation, drying biases, and random moisture losses upon grinding [6].

7.2.2.3 Packing Error

Packing variation can be accommodated by compression (averaging) multiple sample aliquots, by generating a calibration equation on the compressed data and then by predicting duplicate packs of unknown samples [7]. Spinning or rotating sample cups also can reduce this error. The concept is to produce as large a population of measurements as possible noting that the mean of this set of measurements more closely approximates a less “noisy” measurement value. The combination of these methods can reduce prediction errors by as much as 70% (relative) for predicted values as compared to less careful sample collection and presentation.

7.2.3 THE MEANING OF OUTLIERS

Outlier prediction is important during the calibration modeling and monitoring phases. True spectral outliers are considered to be samples whose spectral characteristics are not represented within a specified sample set. Outliers are not considered to be part of the group that is designed to be used as a calibration set. The criterion often given representing outlier selection is a sample spectrum with a distance of greater than three Mahalanobis distances from the centroid of the data. Another quick definition is a sample where the absolute residual value is greater than three to four standard deviation from the mean residual value [8]. Standard multivariate or regression texts more extensively describe these selection criteria.

In a practical sense, outliers are those samples that have unique character so as to make them recognizably (statistically) different from a designated sample population. Evaluation criteria for selecting outliers are often subjective; therefore there is a requirement that some expertise in multivariate methods be employed prior to discarding any samples from a calibration set.

TABLE 7.1
Calibration Error Sources with Recommended Action for Error Reduction

Variance source	Recommended solutions
Nonhomogeneity of sample	<ul style="list-style-type: none"> • Improve mixing guidelines • Improve grinding procedures • Average replicate repacks • Rotate sample cup • Measure multiple readings of large sample volume
Laboratory error	<ul style="list-style-type: none"> • Laboratory audit to correct procedural error • Suggest improvements on analytical procedures • Retrain analysts on procedures • Check and recalibrate reagents, equipment, etc.
Physical variation in sample	<ul style="list-style-type: none"> • Improve sample mixing during sample preparation • Diffuse light before it strikes the sample using a light diffusing plate • Pulverize sample to less than 40-μ particle size • Average multiple repacks • Rotate sample, or average five sample measurements
Chemical variation in sample with time	<ul style="list-style-type: none"> • Freeze-dry sample for storage and measurement • Immediate data collection and analysis following sample preparation • Identification of kinetics of chemical change and avoidance of rapidly changing spectral regions
Population sampling error	<ul style="list-style-type: none"> • Review calibration set selection criteria • Use sample selection techniques such as SUBSET or PICKS used for Selecting Calibration Set
Non-Beer's law relationship (nonlinearity)	<ul style="list-style-type: none"> • Use smaller concentration ranges for each calibration • Use baseline correction such as standard normal variate or polynomial baseline correction • Use one or more indicator variables • Try shorter path lengths • Check dynamic range of instrument
Spectroscopy does not equal manual chemistry	<ul style="list-style-type: none"> • Use different chemical procedures (possibly spectroscopic) • Redefine analytical requirements in terms of known chemistries
Instrument noise	<ul style="list-style-type: none"> • Check instrument performance (i.e., gain, lamp voltage, warm-up time, etc.) • Determine signal-to-noise • Check precision with standard sample replicate measurements
Integrated circuit problem	<ul style="list-style-type: none"> • Replace faulty components
Optical polarization	<ul style="list-style-type: none"> • Use depolarizing elements
Sample presentation extremely variable	<ul style="list-style-type: none"> • Improve sample presentation methods • Investigate wide variety of commercially available sample presentation equipment

(Continued)

TABLE 7.1
(Continued)

Variance source	Recommended solutions
Calibration modeling incorrect	<ul style="list-style-type: none"> • Select and test calibration model carefully • Calculate new equation
Poor calibration transfer	<ul style="list-style-type: none"> • Select calibrations with lowest noise, wavelength shift sensitivity, and offset sensitivity • Identify and transfer actual (not nominal) wavelengths and corresponding regression coefficients
Outlier samples within calibration set	<ul style="list-style-type: none"> • Cumulative normal plots • CENTER program by ISI • DISCRIM by Bran and Luebbe
Transcription errors	<ul style="list-style-type: none"> • Triple-check all handscribed data

7.3 THE CALIBRATION PROCESS

7.3.1 SELECTING THE CALIBRATION SET

The selection or preparation of a set of calibration standards involves following important considerations. The analyst must prepare samples that include the complete range of component concentration as evenly distributed as possible. Random sample selection will cause the mathematical model to most closely fit to the middle concentration samples while the few samples at extremely high or low concentration levels will influence the slope and intercept inordinately. Even concentration distribution will allow the model to minimize the residuals at the extremes and at the center with relatively equal weighting. *Note:* In the past, weighted regression techniques have not received widespread use in NIR; thus the general case for the most commonly used regression algorithms is discussed here.

Besides sample distribution, another critical consideration in selecting/preparing standards is uniform matrix distribution. What is mean here is that the variance in background composition can bring about substantial changes in spectra such as band broadening and peak position shifts. Background characteristics must then be carefully considered when composing the standard calibration set. For example, if the polarity of the selected samples changes with the concentration of various matrix components (e.g., pH) and a polarity change brings about band shifts in the spectra, in order to develop a mathematical model for this application a wide range of pH differences must be included in the calibration standards. A calibration model is then developed representing the types of samples that will be analyzed in the general analysis procedure. Another example often found with solid samples is the effect of moisture content within a solid or powdered sample. The presence or absence of water in a sample will, of course, influence the extent of hydrogen bonding within the sample. Hydrogen bonding will affect both band position and width. If a mathematical model is developed on standards that include a wide range of the component of interest but a small range in moisture, the calibration model will only be useful for samples with a narrow moisture range such as was represented by the standards. Each calibration problem represents something slightly different, yet successful calibration can be performed by paying attention to the finer details.

Other common error sources that are included are technician error in sample preparation, temperature differences in standards or instrument while taking data, calibration standard instability, instrument noise and drift, changes in instrument wavelength setting, nonlinearity, stray light effects, particle size differences, color differences with concentration, solvent interaction differences with change in concentration, and the reference method does not measure the same component as a spectroscopic method. An example of the last problem would typically be found if measuring the percent

composition of protein in a biological solid or liquid. A thermal analysis or Kjeldahl procedure might be used as a primary method to provide reference values to use for calibration. Both thermal analysis and Kjeldahl procedures produce measurements of total reduced nitrogen content. A NIR spectrum does include information on peptide bonds directly, but not on reduced nitrogen. In this type of application the spectroscopic data never will agree perfectly with the spectroscopic data.

When composing a calibration set using volumetric or gravimetric techniques, there are several important items to consider. Often the refractive index of liquids changes with concentration. Particularly problematic are high concentration standards. This problem can be overcome by operating in narrow ranges that approximate linearity, a term sometimes called *range splitting*. The actual absorptivity of a mixture is dependent on the bandpass of an instrument. By leaving the instrument settings of bandpass, scan speed, and response fixed during analysis, this problem can be minimized. The narrower the absorption band of the sample as compared to the bandwidth of the instrument the more deviation from Beer's law occurs. This is due to the fact that a smaller percentage of the total energy passing through or into the sample is attenuated by the absorbing substance. In other words, the narrow absorption band is receiving a greater proportion of light that it does not absorb. Broad bands measured with relatively narrow bandpass instruments do not suffer this problem.

Synthetic samples are generally not considered useful for developing calibration models due to several of the previously mentioned considerations. A rule of thumb for synthetic samples dictates that if there is any spectroscopically observable interaction between the components in the mixture, it will be impractical to produce a standard calibration set. If the only option for the analyst is to produce synthetic standard samples, some rules of mixture design must be followed. In composing standards it has been an unofficial rule of thumb that ten or more samples per term (wavelength or independent variable) should be used in the mathematical model or equation. Actually, having as many samples as required to represent variability of the samples to be measured is a technical requirement. Experimental design can assist the user in composing a calibration set when the total components or variables within the sample have been estimated. This can be done using principal component analysis [9] or by surveying the applications literature for the experience of the previous researchers. For powdered samples, particle size and moisture (O–H stretch) compose the first two major components. Mixture design concepts can be useful in understanding the calibration model concept.

Selecting a sample set for calibration involves the logic and understanding used by pollsters to predict elections. The calibration experiment, however, involves less confusing variables than election polls and much greater confidence is placed on these analytical calibrations. For example, there is generally a substantial a priori knowledge of the chemistry involved in the analytical samples prior to the calibration. Other variables, such as temperature, instrumental variations, and the like, are at least understood in principle. In the election case, war, scandal, economics, or revolutions create unpredictable results. Yet polling science can be useful in selecting calibration samples from large, open populations. If a calibration set is properly selected, the resulting calibration will accurately predict unknown samples.

Calibration sets must not only uniformly cover an entire constituent range; they must also be composed of a uniformly distributed set of sample types. Ideal calibration sets are composed of equal to or greater than 10 to 15 samples per analytical term. These samples ideally have widely varying composition evenly distributed across the calibration range. The sample set must also not exhibit intercorrelation between moisture or particle size with one of the components of interest. An exception to this intercorrelation rule is allowed in obvious cases such as total solids determination or some such application where moisture is an indirect measurement of the total solids. In another well-known case, the hardness of wheat is directly related to the particle size distribution of ground wheat due to the grinder–wheat interaction.

The variance of the sample set used to produce NIR spectroscopy calibration equations determines both robustness and accuracy for a particular application. A calibration learning set that includes a wide variation of sample types and a large constituent range will allow a calibration

model where a wider range of materials may be analyzed, but with a resultant loss in accuracy. If the calibration learning set has a small variance in sample type and a narrow constituent range, the accuracy for analyzed samples within the range is increased, but fewer unusual samples can be analyzed with confidence using this approach. Thus, for quality control procedures, one may wish to have both calibrations available. The robust calibration is used to detect “outliers,” while the dedicated calibration is used to accurately measure the constituent values of “normal” samples. Outlier detection techniques can be used to predict the “uniqueness” of a sample using the H statistic, also known as Mahalanobis distance.

7.3.2 DETERMINATION OF STANDARD CONCENTRATIONS: MANUAL METHODS

Finding the composition of standard samples is not a problem when using synthetic mixtures for calibration. Samples that are not synthetic and that are characterized as complex in nature must be analyzed using an accepted or authorized chemical procedure prior to calibration. Several error sources are introduced into the calibration data that are produced by the primary analytical method. For example, if the analyst were working with a liquid in order to characterize moisture content using spectroscopy, an oven or Karl Fischer titrimetric method would be selected to determine the moisture content in the standard set. Following this procedure, a calibration model would be developed using the primary method composition and relating it to the spectroscopic data. For each technique used as a primary method there are inherent errors in measurement. This error must be considered when developing the calibration equation model. Random error within the chemical data must be reduced to the minimum possible level in order to perform the optimum calibration. Error in the chemical (laboratory data) cannot be removed mathematically but it can be reduced by using the average composition as determined by replicate chemical measurements for each sample.

7.3.3 COLLECTING OPTICAL DATA

Collecting optical data using a set of standards involves several noteworthy points. The analyst must be aware that the instrument varies from day to day, and within a single day. Thus samples measured on one day may not agree with samples measured on another day. An instrument reading will vary on the same day with lamp temperature and usage. As the temperature of the lamp and detector equilibrate, the instrument reaches its maximum stability. Until there is no temperature gradient in the instrument, it will drift. The user should be alert to such problems as placing instrumentation near air conditioner or heating vents, as the instrument may drift depending on changes in temperature and humidity. Note that humidity also plays an important role in any infrared (IR) instrument. Measuring standards when low relative humidity is in the environment will not allow accurate prediction of samples in high humidity. Changes in spectra may be due to both sample changes and increased absorption of energy due to moisture within the instrument.

Data should be collected by random sample selection for both calibration standards and, if possible, routine samples. Spectroscopic measurement should be performed under as identical conditions as possible between calibration conditions and routine operations. Data should not be collected in order of concentration or the mathematical calibration model might correlate to changes in an instrument that have no relationship to changes in the concentration of samples.

For routine measurements of complex samples for which the instrument will be calibrated for relatively long periods, it is advisable to measure calibration standards over a several day period and under varying temperature conditions. In this case the instrumental variation from day to day will be “ignored” or minimized in the resulting mathematical calibration model.

For routine, single-component measurements these precautions seem to be overly rigorous. However, multicomponent problems are inherently more complex and require the proper management of several variables in order to develop a suitable calibration. The ultimate goal of calibration is

to calculate a mathematical model of the calibration data that is most sensitive to changes in concentration of the sample and least sensitive to nonconcentration-related factors, such as physical, chemical, and instrumental variables. Every case must be evaluated in terms of the chemical and physical data of samples and the information that the analyst wishes to obtain. For example, specular reflectance of surfaces contains information for refractive index, surface film thickness, and the bulk optical properties of a sample. Scattering from particles is wavelength-dependent, with the high-frequency/high-energy light subject to the most scattering and the low-frequency/low-energy light subject to the least amount of scattering; the higher the frequency, the greater the sample penetration. In highly scattering samples, NIR energy may penetrate only to one or two particle thicknesses in depth. With highly absorbing samples, penetration may be up to several millimeters.

When highly crystalline materials are measured, the analyst may use a ground-glass diffusing element, or Fresnel lens between the source and the sample. This type of approach can bring about decreased sensitivity of the calibration to surface packing of the sample. The use of polarized light in diffuse reflectance of heavily oriented crystals may prove useful. The common method used to decrease sensitivity to scattering samples is to pulverize and thoroughly mix them, and to rotate the sample cup during measurement. These considerations can be helpful when making measurements of highly scattering samples. It should be noted that monochromators employing diffraction gratings rather than dispersion prisms produce increased p-polarized light over the s-polarized component. Instruments using different monochromator types and sampling geometries will ultimately produce slightly different reflectance data. Bandpass, response, and scan speed also affect the actual raw reflectance data and should be noted when reporting such data in technical reports and scientific journals.

The effects of changing any of these instrument parameters during calibration and routine analysis can be disastrous. The analyst must establish the measurement and instrument environment conditions and maintain these throughout the quantitative calibration and prediction (analyses) work. The problems associated with differences in the instrument environment have been studied while using mobile van laboratories. Temperature variation alone between summer and winter conditions has brought significant challenges to accurate and reproducible results [4].

7.3.4 DEVELOPING THE MATHEMATICAL MODEL

Mathematical modeling of calibration standards has been addressed extensively both in this book and in the general literature. It is the intention of this chapter to point the analyst to several literature sources for detailed information while highlighting some of the more important principles of mathematical calibration modeling techniques. It is the best mathematical model that when applied to photometric data obtained from a sample of unknown concentration will compute a reliable, accurate, and reproducible value of percent composition for the component(s) of interest. The ideal mathematical model will be “blind” to endogenous instrument variation, temperature effects, background interferences, and the like, and at the same time be most sensitive to changes in the concentration of the component(s) of interest.

The operation of NIR spectroscopy instruments involves measurement of the reflected incident beam from the sample surface at a number of wavelengths and from a standard reflecting surface at the same wavelengths. Thus, in practice the sample reflectance measurement is a relative measurement as compared to a nonstandard reflector. The relationship:

$$\text{Reflectance} = \frac{\text{Intensity of sample reflectance}}{\text{Intensity of reference reflectance}}$$

exists at every wavelength measured. The K–M function, which includes formalized scattering and absorption terms, is not generally used in NIR spectroscopy, except for particle size measurements or for highly scattering solid samples. Logarithmic reflectance terms are most often used in deriving

a calibration equation. The scattering effects of the sample are somewhat compensated for in the calibration model most often by using first or second derivative math pretreatments to partially compensate for baseline offset between samples and to reduce instrument drift effects. Quite often the most important use of derivatives is to glean additional information from the spectrum — information that is not resolved using $\log(1/R)$ only.

The general statistical expression used for NIR spectroscopy calibrations is a multiple linear regression (MLR) type. In practice, the regression equation is solved for the best set of coefficients that when combined with the spectral data predict the constituent content for unknown samples. It is always possible to solve the regression equation for a single set of calibration samples (closed set), although the purpose of the calibration is to obtain a general or “universal” equation that allows for accurate prediction of constituent values in any sample the analyst might encounter (Figure 7.2).

Our discussion up to this point has brought forward many of the important factors of experimental design that must be considered when formulating the optimum calibration. Proper setup of calibration sets yields calibrations that are superior in terms of both precision and accuracy. The “garbage-in, garbage-out” adage applies in NIR spectroscopy. The computer is used simply to perform repetitive calculations faster and more accurately than are possible with manually computed results. The complexity of calibration development has not allowed completely automated mathematical modeling in any current software. In any case, the most critical step in calibration is the proper collection of samples to represent the population for routine analysis.

Several models have been proposed for automatic calibration. These include spectral matching of unknown samples to standards that have already had a calibration developed for them. The procedure would in effect “select” the proper calibration model for the samples presented. Other proposed autocalibration methods have included ideas where a product is identified as to type and then the selection of a calibration equation is based on the estimated range of the constituent. Thus, a different calibration line would be used for material A at 10 to 15% analyte concentration than would be used for the same material at, say, 15 to 20% analyte. This technique has been termed *product identification*, followed by range splitting or “bracketing.” Other more sophisticated approaches involve using discriminant analysis to sort samples by spectra or equation type and then applying the selected equation to the sample data. All of these approaches are interesting and provide plenty of opportunity for additional work in the area of calibration techniques, yet the commonly used calibration techniques include the rigor of experimental design, calibration modeling, and validation testing. Short-cuts to these methods are experimental, theoretical, and when attempted by the uninitiated will result in great potential for failure.

TABLE 7.2
Description of Independent Variable Terms (Math Pretreatment or Data Decomposition Techniques)

Name of ind. var. term	Equivalence in $-\log R$ terms
$-\log R$	1
First difference “derivative”	2
Second difference “derivative”	3
Higher order differences “derivatives”	≥ 4
Difference ratios	≥ 4
Fourier coefficients	1 — full spectrum
Principal components scores	1 — full spectrum
Factor analysis scores	1 — full spectrum
Curve-fitting coefficients	1 — full spectrum
Partial least-squares scores	1 — full spectrum

So how does one select proper mathematics to calibrate for the samples at hand? Table 7.2 demonstrates the common algorithms for multivariate calibration. Note that first- through fourth-order functions are used with either zero or natural y intercepts. The method of least squares is used to obtain these calibration equations. When a small number of standards are used or when the relationship between the optical data and concentration is not known, a segment fit can be used. The segment fit assumes nothing about the data other than the fact that each datum has the same significance in the calibration and that predicted values are based only on a series of lines that connect the data points. A segment fit produces no goodness-of-fit statistics due to the fact that it always forms a “perfect” fit to the data. It can be useful in single-component rapid analysis or feasibility studies. A complete univariate software package will include segment fitting as a convenience for analysts performing “quick-and-dirty” experiments, but this type of calibration is not particularly useful for serious analyses. Linear calibrations, also referred to as a first-order or univariate case, are most commonly used for simple calibrations of a single component with the relationship for concentration vs. optical data assumed to follow Beer’s law. Multicomponent methods, also termed *multivariate techniques*, are used when a single wavelength will not predict concentration due to band overlap or when more than one component is to be measured simultaneously.

Simple multicomponent measurements are sometimes used when the spectra of two or more chromophores within a mixture are found to be overlapped across the entire spectral region of interest. When performing multivariate analysis on such mixtures, the analyst must select wavelengths where the molar absorptivities for the components are most different. With complex matrices, it is impossible to know if an ideal set of wavelengths has been selected for a calibration equation. Often when new instrumental effects are indicated and when new calibration sets are selected, new wavelengths are chosen for the same application. It has been somewhat dissatisfying to be caught in a trap of juggling wavelengths always hoping to find the best combinations for a particular analytical task. In multilinear regression, the analyst assumes that adding new or different wavelengths will provide a calibration equation compensating for noise, interferences, nonlinearity, background scattering, and other “nasty” deviations from Beer’s law; sometimes this is true and sometimes not [10–15,33].

PCR has been applied in a variety of calibration techniques with reported success [16–20]. This technique has the obvious benefit of not requiring wavelength searches. The analyst must decide how many components (or eigenvectors) to use in the calibration equation. PLS is another multivariate technique that seems to show promise in calibrations where samples are most complex and where low signal-to-noise relationships exist between analyte and spectral features. PLS has wide-ranging applications in a variety of spectroscopic and scientific situations [21–27]. PLS also shows promise in color modeling. Table 7.2 illustrates the most common mathematical pretreatments and their equivalence in $\log(1/R)$ terms. Table 7.3 gives several arguments for and against the various calibration modeling techniques.

7.3.5 VALIDATING THE MATHEMATICAL MODEL

The following tests (Table 7.4) are used to determine NIR spectroscopy equation acceptance:

1. Standard Methods
 - (a) Coefficient of determination
 - (b) Correlation coefficient
 - (c) F -test statistic (F for regression)
 - (d) Confidence limits for prediction values based on F
 - (e) Partial F or t^2 test for a regression coefficient
 - (f) Standard error of estimate (standard error of calibration)
 - (g) Bias-corrected standard error
 - (h) Standard error of cross-validation
 - (i) Standard error of prediction
 - (j) Standard error of laboratory

TABLE 7.3
Using a Variety of Mathematical Modeling Techniques

Technique	Comments
MLR	A: Most robust for all math pretreatments A: Mathematical properties explored and best understood A: Spectral interpretation most straightforward
$-\log R$	A: Eliminates truncation errors due to additional data transformation A: Easiest to use and interpret A: Proper regression technique will compensate for additive and multiplicative error
Difference math	A: Will resolve band overlap A: Reduction of additive (offset) error A: Compensates for instrument drift A: Enhances definition of difficult or "hidden" spectral information A: Valuable as a visual research tool for spectroscopists A: May show predictive advantage for spectra with large offset or difficult-to-resolve data D: Will show apparent improvement on calibration statistics; could result in overfit of the calibration data D: Far more sensitive to instrument noise, thus smoothing required D: Constrains the search algorithm and prevents optimal coefficient selection D: Complicates spectral interpretation empirically
Difference ratios	A: Apparent calibration advantage as determined empirically A: May find applications in online work where variation in the baseline may present problems
FT-NIR	A: Can be used to "filter" high-frequency noise from spectral data A: Allows 95% compression of data for storage A: Allows convolution and deconvolution for band identification; No distinct advantage for prediction
PCA	A: One solution to the wavelength problem A: Can be used on fixed-filter instruments to enhance calibration A: Allows greater capability in spectral matching (search), pure component delineation, noise filtering, and discriminant analysis D: Difficult to interpret spectrally; Some predictive improvement in low signal-to-noise analyte problems
PLS	A: Gives weighted PCR calibration (added dimensionality) A: All the advantages of PCR A: Can be more robust than MLR A: Some predictive improvement for difficult constituents with low signal-to-noise/low detection limits D: User must be particularly cautious to avoid overfitting of calibration spectra to include instrumental artifacts D: In general, it is more difficult to transfer calibrations between instruments
Factor analysis (FA)	A more generic term for PCA PLS-type algorithms
Curve fitting	A technique for identifying individual component bands from overlapping spectra; useful as a spectral evaluation technique

A = advantage; D = disadvantage.

TABLE 7.4
Statistical Tests Common to NIR Methods (Common Symbols)

R^2	coefficient of multiple determination
Σ	capital sigma represents summation or sum of all values within parenthesis
N	notation for total number of samples used for computation
y_i	a singular y value for the i th sample. In real cases it represents all the actual y values for the 1st through N th data points
\hat{y}_i	symbol for the estimated y value given a regression line. Therefore for any given set of x values, for an i th sample there is a corresponding \hat{y} or estimated value based on the x_i values. It can also be stated as an analytical value derived by the calibration equation
\bar{y}	mean y values for all samples where y is some function of x as $y = f(x)$
K	number of wavelengths used in an equation
r	simple correlation coefficient for a linear regression for any set of data points; this is equal to the square root of the R^2 value
b_0	bias or y-intercept value for any calibration function fit to x, y data. For bias corrected standard error calculations the bias is equal to the difference between the average primary wet chemical analytical values and the NIR-predicted values
B_i	regression coefficient at one wavelength used in a regression equation

- (k) Standard deviation of difference (SDD)
- (l) Offset sensitivity (index of systematic variation)
- (m) Random variation sensitivity (index of random variation)
- 2. Methods for Testing PCR and PLS Equations
 - (n) Prediction sum of squares (PRESS)
 - (o) Cross-validation
 - (p) Validation plots

Explanations for these statistical tests can be found in standard statistical texts and instrument manufacturers' instruction manuals. Detailed description of these methods are beyond the scope of this chapter.

7.3.6 ROUTINE ANALYSIS AND MODELING

Statistics useful in monitoring the validity of a calibration equation as well as the fitness of individual samples to calibration include:

1. Lab and NIR means differences
2. Bias control limits
3. Unexplained control limits
4. Lab and NIR standard deviation differences
5. Slope
6. Coefficient of determination (R^2)

All of the previously mentioned statistics are useful in determining both the feasibility of an NIR method as well as calibration equation validity for well-known applications. The statistics allow us to determine the fitness of a particular calibration; they allow us to select the proper numbers and locations of wavelength choices using purely statistical criteria; and they give us some indication of the ability of a particular calibration to accurately predict analyte concentrations in samples never previously presented to the instrument. The use of bias and slope adjustments to improve calibration or prediction statistics requires some precautionary statements. As illustrated in Figure 7.2, if monitor

sample sets are accurately analyzed, routine analysis is assumed to be validated. If, on the other hand, a calibration model fails during the validation or monitoring steps, a complete recalibration is the most correct procedure. Bias and slope corrections are used to adjust a calibration model in order to correct differences in wet laboratory methods/procedures between the calibration sample set and monitor sample sets. Prediction errors requiring continued bias and slope corrections indicate drift in laboratory chemistry or instrument photometric or wavelength instability.

7.3.7 TRANSFER OF CALIBRATIONS

Technically valid calibrations transfer is not the trivial process that some would propose. In fact, due to advancing calibration mathematics such as PCR, PLS, and spectral matching/search algorithms, it becomes even more critical that transfer technologies be scientifically scrutinized. To date, the most successful approach for transferring calibrations for use with all multivariate mathematical modeling techniques is found in a three-pronged approach:

1. Make instrumentation more alike and more reliable. This requires that center wavelength, bandpass, response, wavelength accuracy, and reproducibility be within close tolerances between instruments where transfer is to be attempted.
2. Relatively small differences between instruments are corrected using firmware or computer software to make the instruments virtually identical.
3. Calibrations to be used for transfer are developed to be insensitive to temperature and instrument variation to within specified tolerances.

When these procedures are correctly implemented, calibration transfer is routine and valid. Of course, all calibrations do not require this rigor, but these techniques are appropriate as a “universal” procedure.

7.3.8 PERFORMING AN ACTUAL FEASIBILITY STUDY

In determining whether or not a particular chemical component can be measured using NIR spectroscopy, it is helpful to run through a checklist of important steps:

1. Determine whether transmission or reflectance is most appropriate.
2. Develop a sample presentation procedure that is both rapid and reproducible.
3. Measure spectra for high, medium, and low concentration levels.
4. Measure several spectra for each of the samples, taking different aliquots for each in order to determine repeatability and estimate detection limit.

A comparison of these results will allow a preliminary judgment as to whether more samples should be measured to determine more accurate estimates for SEP and SDD (Table 7.4 and Table 7.5).

7.4 STATISTICAL TERMS

Sum of squares regression (SS_{regr}):

$$SS_{\text{TOT}} - SS_{\text{res}} = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$$

Sum of squares for residuals (SS_{res}):

$$SS_{\text{res}} = \sum_{i=1}^N (y_i - \hat{y})^2$$

Mean square for regression (MS_{regr}):

$$\frac{SS_{\text{TOT}} - SS_{\text{res}}}{\text{d.f. for regr.}} = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{K + 1}$$

Mean square for residuals (MS_{res}):

$$\frac{SS_{\text{res}}}{\text{d.f. for residual}} = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - K - 1}$$

Total sum of squares (SS_{TOT}):

$$SS_{\text{TOT}} = \sum_{i=1}^N (y_i - \bar{y})^2$$

7.5 TEST STATISTICS

Statistics: F -Test Statistic for the Regression

Abbreviations: F , T^2

Summation Notation:

$$F = \frac{\left[1 - \left(\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / (N - K - 1)}{\sum_{i=1}^N (y_i - \bar{y})^2 / (N - 1)} \right) \right] (N - K - 1)}{\left[1 - \left(\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / (N - K - 1)}{\sum_{i=1}^N (y_i - \bar{y})^2 / (N - 1)} \right) \right] (K)}$$

TABLE 7.5
Simple Procedure to Develop a Feasibility Calibration

1. Collect approximately 30 to 50 samples covering the entire concentration range to the constituent of interest
 2. The conventional chemical analyses on these samples must be as accurate as possible. The chemical values used must be the mean results for blind duplicate analyses using an official chemical procedure
 3. There should be no intercorrelation between major constituents unless this relationship always exists within this type of material
 4. Eventual variations in particle size, sample presentation, and process conditions must be represented within the test sample population
 5. The calibration model is tested using cross-validation methods such as SECV and PRESS (see Table 7.4)
 6. Overall quality of the calibration is tested using the statistical methods found within Table 7.4
 7. If wavelength calibrations are used, the selected wavelengths utilized in the calibration equation should match the vibrational absorbance frequencies known to correspond to the constituent of interest
 8. A validation sample set should be used to determine approximate SEP value
-

Computational Formula:

$$F = \frac{R^2(N - K - 1)}{(1 - R^2)(K)} = \frac{MS_{\text{regr}}}{MS_{\text{res}}}$$

Comments: Also termed F from regression, or t^2 . F increases as the equation begins to model, or fit, more of the variation within the data. With R^2 held constant, the F value increases as the number of samples increases. As the number of wavelengths used within the regression equation decreases, F tends to increase. Deleting an unimportant wavelength from an equation will cause the F for regression to increase.

The F statistic can also be useful in recognizing suspected outliers within a calibration sample set; if the F value decreases when a sample is deleted, the sample was not an outlier. This situation is the result of the sample not affecting the overall fit of the calibration line to the data while at the same time decreasing the number of samples (N). Conversely, if deleting a single sample increases the overall F for regression, the sample is considered a suspected outlier. F is defined as the mean square for regression divided by the mean square for residual (see statistical terms in this table).

Statistic: Coefficient of Multiple Determination

Abbreviations: R^2 or r^2

Summation Notation:

$$R^2 = 1 - \left(\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / (N - K - 1)}{\sum_{i=1}^N (y_i - \bar{y})^2 / (N - 1)} \right); \quad R^2 = \left(\frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \right)$$

Adjusted R^2 for proper degrees of freedom; unadjusted R^2 (general notation).

Computational Formula:

$$R^2 = 1 - \frac{SEC^2}{(SD_{\text{range}})^2} = \frac{SS_{\text{TOT}} - SS_{\text{res}}}{SS_{\text{TOT}}}$$

Comments: Also termed the R^2 statistic, or total explained variation. This statistic allows us to determine the amount of variation in the data that is adequately modeled by the calibration equation as a total fraction of 1.0. Thus $R^2 = 1.00$ indicates the calibration equation models 100% of the variation within the data. An $R^2 = 0.50$ indicates that 50% of the variation in the difference between the actual values for the data points and the predicted or estimated values for these points are explained by the calibration equation (mathematical model), and 50% is not explained. R^2 values approaching 1.0 are attempted when developing calibration. R^2 can be estimated using a simple method as outlined below.

$$R_{\text{EST}}^2 \approx 1 - \frac{SEL^2}{SD_{\text{range}}^2}$$

where SEL = standard error (deviation) of the wet laboratory, and

SD_{range} = standard error (deviation) of the x values (concentration values)

Statistic: Confidence Limits for Predicted Values

Abbreviation(s): $\hat{Y} \pm \text{C.L.}$

Summation Notation:

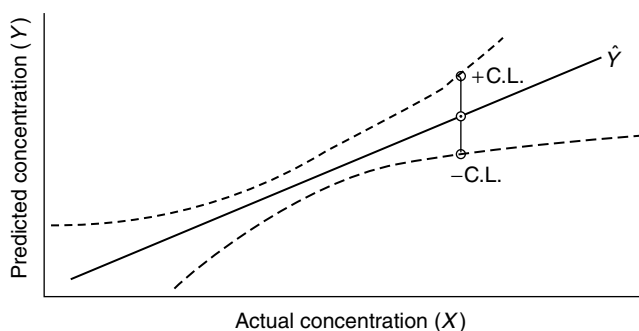
$$\text{Confidence interval} = \hat{Y} \pm (2F)^{1/2} \left\{ \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - K - 1} \left[\frac{1}{N} + \frac{(x_i - (\sum_{i=1}^N x_i/N))^2}{\sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i/N)^2} \right] \right\}^{1/2}$$

where F is $F(2, N - K - 1, 1 - \alpha)$ from the F -distribution table.

Computational Formula:

$$\text{Confidence interval} = \hat{Y} \pm (2F)^{1/2} (\text{SEC}) \left[\frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \right]^{1/2}$$

Comments: Also termed confidence bands, confidence intervals for a calibration line, or 95%* confidence limits for a calibration line. We note that the confidence interval for any statistic is equal to the statistic $\pm t$ or the square root of $2F$ times the standard error of the statistic. In calibration, the confidence limit allows us to estimate the precision of the analytical method for samples of differing actual concentrations. Often confidence limits can be shown as follows.



This relationship can be calculated for most X, Y situations where $\hat{Y} = f(x)$.

Statistic: Student's t value (for a regression)

Abbreviation(s): $t, \sqrt{F}, F^{1/2}$

Summation Notation: See coefficient of multiple determination for summation notation of R^2 .

Computational Formula:

$$t = \frac{R(N - K - 1)^{1/2}}{(1 - R^2)^{1/2}} = F^{1/2} = \left\{ \frac{\text{MS}_{\text{regr}}}{\text{MS}_{\text{res}}} \right\}^{1/2}$$

t test for regression.

Comments: This statistic is equivalent to the F statistic in the determination of the correlation between X and Y data. It can be used to determine whether there is a true correlation between an NIR value and the primary chemical analysis for that sample. It is used to test the hypothesis that the correlation really exists and has not happened only by chance. A large t value (generally greater than 10) indicates a real (statistically significant) correlation between X and Y .

Statistic: Student's t test for the residual or the difference between the NIR analysis and the primary wet chemical analysis for any sample.

Abbreviations: $t, F^{1/2}$

Summation Notation: See SEC summation notation.

Computational Formula:

$$t = \frac{\text{Residual}}{\text{SEC}} = \frac{(y_i - \hat{y}_i)}{\text{SEC}}$$

Comments: This test allows evaluation criteria for assessing the variation between an NIR value and its primary chemical value. t values of greater than 2.5 are considered significant and those NIR analyses having such large t values may possibly be outliers. Replicate primate wet chemistry should be performed on such samples. If the wet chemistry is proven correct, the sample may be a spectral outlier. Most often, high t -test values here indicate poor laboratory results or a problem with sample presentation, such as grinding, or packing of the sample cup.

Statistic: Partial F or t^2 test for a regression coefficient

Abbreviation: t^2

Summation Notation: See statistical terms (p. 140).

Computational Formula:

$$\text{Partial } F = \frac{\text{SS}_{\text{res}}(\text{all variables except 1}) - \text{SS}_{\text{res}}(\text{all variables included})}{\text{MS}_{\text{res}}(\text{all variables})}$$

Comments: This test indicates if the addition of a particular wavelength (independent variable) and its corresponding regression coefficient (multiplier) add any significant improvement to an equation's ability to model the data (including the remaining unexplained variation). Small F or t values indicate that no real improvement is given by adding the wavelength into the equation.

If several wavelengths (variables) have low t or F values (less than 10 or 100, respectively), it may be necessary to delete each of the suspect wavelengths, singly or in combination, to determine which wavelengths are the most critical for predicting constituent values. In the case where an important wavelength is masked by intercorrelation with another wavelength, a sharp increase in the partial F will occur when an unimportant wavelength is deleted and where there is no longer high intercorrelation between the variables still within the regression equation.

The t statistic is sometimes referred to as the ratio of the actual regression coefficient for a particular wavelength to the standard deviation of that coefficient. The partial F value described is equal to this t value squared; noting that the t value calculated this way retains the sign of the coefficient, whereas all F values are positive.

Statistic: Standard error of estimate (SEE), also termed standard error of calibration (SEC).

Abbreviations: SEE, SEC

Summation Notation:

$$\text{SEE} = \text{SEC} = \left(\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - K - 1} \right)^{1/2}$$

Computational Formula:

$$\text{SEE} = \text{SEC} = \{\text{MS}_{\text{res}}\}^{1/2}$$

Comments: This statistic is the standard deviation for the residuals due to differences between actual (primary wet laboratory analytical values) and the NIR predicted values for samples *within* the calibration set. It is an indication of the total residual error due to the particular regression equation to which it applies. The SEC will decrease with the number of wavelengths (independent variable

terms) used within an equation, indicating that increasing the number of terms will allow more variation within the data to be explained, or “fitted.” The SEC statistic is a useful estimate of the theoretical “best” accuracy obtainable for a specified set of wavelengths used to develop a calibration equation. The SEE is another term for SEC, both terms designating the square root of the mean square for residuals. In this calculation, the residual for each sample is equal to the actual chemical value minus the NIR predicted value for all samples *within* the calibration set.

Statistic: The Standard Error of Prediction

Abbreviation(s): SEP

Summation Notation: See statistical terms (p. 140).

Computational Formula:

$$\text{SEP} = \{\text{MS}_{\text{res}}\}^{1/2}$$

Comments: The SEP is also termed the standard error of performance, also known as the standard deviation for the residuals due to differences between actual (primary wet chemical analytical values) and the NIR predicted values for samples *outside* of the calibration set using a specific calibration equation.

The SEP is calculated as the root mean square differences (RMSD), also known as a mean square for residuals for $N - 1$ degrees of freedom. It allows for comparison between NIR-observed predicted values and wet laboratory values. The SEP is generally greater than the SEC but could be smaller in some cases due to chance alone. When calculating the SEP, it is critical that the constituent distribution be uniform and the wet chemistry be very accurate for the validation sample set. If these criteria are not met for validation sample sets, the calculated SEP may not have validity as an accurate indicator of overall calibration performance. To summarize, the SEP is the square root of the mean square for residuals for $N - 1$ degrees of freedom, where the residual equals actual minus predicted for samples *outside* the calibration set.

Statistic: Standard Error of Cross-Validation (SECV)

Abbreviation(s): SECV

Summation Notation: See statistical terms (p. 140).

Computational Formula:

$$\text{SECV} = \{\text{MS}_{\text{res}}\}^{1/2}$$

Comments: The calculation of SECV is a method for determining the “best” number of independent variables to use in building a calibration equation. The SECV method is based on an iterative (repetitive) algorithm that selects samples from a sample set population to develop the calibration equation and then predicts on the remaining unselected samples. Some procedures for calculating SECV may calibrate using two-thirds of the samples while predicting on the remaining one-third of the samples. The SECV is an estimate of the SEP and is calculated as SEP or SECV as the square root of the mean square of the residuals for $N - 1$ degrees of freedom, where the residual equals the actual minus the predicted value. In most statistical software packages, the SECV is computed for several different equations, the equation with the lowest SECV being selected as the best calibration.

This method for testing calibration models is often used for MLR, PCR, and PLS.

Statistics: The Bias-Corrected Standard Error

Abbreviation(s): SEC(C), SEP(C)

Summation Notation: See statistical terms (p. 140).

Computational Formula:

$$\text{SEC(C)} = \text{SEP(C)} = \left(\frac{\sum_{i=1}^N (y_i - \hat{y}_i + b_0)^2}{N - 1} \right)^{1/2}$$

Comments: Bias-corrected standard error measurements allow the characterization of the variance attributable to random unexplained error. The bias value is calculated as the mean difference between two columns of data, most commonly actual minus NIR predicted values.

Statistic: SDD

Abbreviations: SDD, SED_{duplicates}

Summation Notation: See computational formula as follows.

Computational Formula:

$$\text{SDD} = \text{SED}_{\text{duplicates}} = \left\{ \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N} \right\}^{1/2} = \left\{ \frac{\sum (y_1 - y_2)^2}{2N} \right\}^{1/2}$$

where y_i is an individual analytical result (one of several replicate analyses)

\bar{y} is the average analytical results for all replicate values

N is total number of samples

y_1 = first analysis

y_2 = second analysis

Comments: The repack error, sometimes referred to as the SDD or standard error of differences for replicate measurements (SED replicates), is calculated to allow accurate estimation of the variation in an analytical method due to both sampling and presentation errors. It is a measure of precision for an analytical method.

Statistic: Offset Sensitivity

Abbreviation(s): OS, ISV

Summation Notation: See computational formula as follows.

Computational Formula:

$$\text{OS} = \text{ISV} = \sum_{i=1}^N \beta_i$$

Comments: Also termed *systematic variation* or *index of systematic variation* (ISV); OS is equal to the sum of all regression coefficients or PCR/PLS factors. The larger the value, the greater is the sensitivity to particle size differences between samples or to the isotropic (mirrorlike) scattering properties of samples. The offset sensitivity is used to compare two or more equations for their “blindness” to offset variation between samples. Equations with large offset sensitivities indicate that particle size variations within a data set may cause wide variations in the analytical result.

Statistic: Random Variation Sensitivity

Abbreviation(s): RVS = {IRV}₂¹

Summation Notation: See computational formula as follows.

Computational Formula:

$$\text{RVS} = (\text{IRV})^{1/2} = \sum_{i=1}^N |\beta_i|$$

Comments: This statistic is also termed the index of random variation (IRV). Random variation sensitivity is calculated as the sum of the absolute values of all regression coefficients, multipliers, or B coefficients (for PLS regression). The larger the value, the greater the sensitivity to factors such as poor wavelength precision, temperature variations within samples and instrument, and electronic noise. The higher the value, the less likely the equation can be transferred successfully to other instruments.

Statistics: Standard Error of the Laboratory (SEL) for Wet Chemical Methods

Abbreviation(s): SEL

Summation Notation: See formulas as follows.

Comments: The SEL can be determined by using a single sample analyzed in replicate by one or more laboratories. The average analytical value for the replicates on a single sample is determined as

$$\bar{y} = \frac{\sum_{i=1}^N y_i}{N} = \frac{y_1 + y_2 + \cdots + y_n}{N}$$

If duplicate analyses are used within a single laboratory, our SEL is given by

$$\text{SEL} = \left\{ \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N} \right\}^{1/2} \equiv \left\{ \frac{\sum_{i=1}^N (y_1 - y_2)^2}{2N} \right\}^{1/2}$$

where y_i = individual laboratory result

\bar{y} = mean of two lab results $(y_1 + y_2)^{1/2}$

N = total number of samples (not total number of analyses)

y_1 = first analysis

y_2 = second analysis

y_N = N th analysis

If multiple replicate analyses are performed (greater than three), we calculate SEL as

$$\text{SEL} = \left\{ \frac{\sum_{j=1}^M \sum_{i=1}^N (y_{ij} - \bar{y}_j)^2}{N(r-1)} \right\}^{1/2}$$

This can apply whether the replicates were performed in a single laboratory or if a collaborative study was undertaken at multiple laboratories. Additional techniques for planning collaborative tests can be found in Reference 6.

Note: r = number of replicate analyses for each sample.

7.5.1 PRINCIPAL COMPONENTS REGRESSION AND PARTIAL LEAST SQUARES TESTING

Statistic: PRESS (Prediction Sum of Squares)

Abbreviation(s): PRESS

Summation Notation: See text as follows.

Comments: PRESS allows the determination of the optimum number of factors, also termed dimensions, *PLS components*, *eigenvectors*, *principal components* or *principal factors*. PRESS is computationally intensive and results in the use of substantial computer time. The procedure would normally start by using one factor as a model while leaving a single sample out of the calibration set. After a calibration is developed using the remaining samples, the algorithm is programmed to predict the one selected sample and record the difference between the actual vs. the predicted values. When this procedure is iterated (repeated) for the entire sample set, the sum of squares for the residuals, $\sum_{i=1}^N (y_i - \hat{y}_i)^2$, is reported. The procedure then adds another factor (now two) and repeats the process. Finally, the PRESS procedure will stop when the predesignated number of factors is reached (say 10 to 20 maximum) and at that time the calibration model with the least number of factors and the smallest sum of squares for residuals is selected as the best model *for the calibration set used*. A plot of PRESS values (y-axis) vs. the number of factors (x-axis) is often used to determine the minium PRESS corresponding with the smallest number of factors in the calibration model. An excellent description of this procedure (algorithm) is found in Reference 8 (p. 325). Note that PRESS results are often presented or reported in two columns as number of factors, and sum of squares for residuals or PRESS values as columns one and two, respectively.

Statistic: Cross-validation

Abbreviation(s): SECV

Summation Notation: See text as follows.

Comments: The cross-validation algorithm is performed identically to PRESS with the exception that rather than sum of squares for residual as the reported results, the cross-validation uses the square root of the mean squares for residuals, using $N - 1$ degrees of freedom for each model as:

$$\text{SECV} = \{\text{MS}_{\text{res}}\}^{1/2} = \left\{ \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - 1} \right\}^{1/2}$$

Thus a cross-validation report would traditionally include two columns; column one as number of factors, and column two as the SECV.

Statistic: Validation by Prediction Using SEP

Summation Notation: See previous discussion for SEP.

Comments: A final calibration model is often tested by predicting a set of samples alike in composition to the calibration set. The SEP is then calculated and compared to the SEL and SEC as to acceptability of performance. If errors are not acceptable, recalibration steps should be reviewed.

An excellent discussion on additional procedures in evaluating PLS modeling is found in Reference 28 (pp. 70–71).

7.5.2 ASTM PRACTICE FOR MULTIVARIATE REGRESSION

An official ASTM (American Society for Testing and Materials, 100 Barr Harbor Dr., West Conshohocken, PA 19428) document, “Standard Practices for Infrared, Multivariate, Quantitative Analyses,” has been published [29]. This 25-page document was first designated E1655-94 and later E1655-97 (revised). The document includes “a guide for the multivariate calibration of IR spectrometers used in determining the physical or chemical characteristics of materials.” The practice applies to “the NIR spectral region (roughly 780 to 2500 nm) through the Mid-Infrared (MIR) spectral region (roughly 4000 to 400 cm^{-1}).” In addition, the practice also includes procedures for collecting and treating data for developing IR calibrations; definitions for terms and calibration techniques; and criteria for validating the performance of a calibration model.

While several official chemometric methods of analysis have published, this practice provides the first official working document for the application of chemometric multivariate analysis techniques

to NIR and IR spectrometers. Technical discussions are included for selecting the calibration set; determination of concentrations or properties, or both, for calibration samples; and the collection of IR (and NIR) spectra. In addition, the topics of calculating the mathematical model, validating the mathematical model, and applying the finished model for analysis of unknowns is described; the technical issues for routine analysis and monitoring, and calibration transfer are included within the practice. The mathematics of MLR, PCR, and PLS are delineated; with a discussion of the estimation of parameter values from spectra also included [29–32].

ACKNOWLEDGMENT

The author is grateful to Academic Press for permission to use sections from H. Mark and J. Workman, Calibration: Developing the Calibration Model, and Calibration: Auxiliary Statistics for the Calibration Model, Chaps. 36 and 37, in *Statistics in Spectroscopy*, Elsevier Science, Amsterdam, 2003.

REFERENCES

1. G. Kortum, *Reflectance Spectroscopy*, Springer-Verlag, New York, 1969, pp. 103–163.
2. J. M. Olinger and P. R. Griffiths, *Anal. Chem.*, **60**: 2427 (1988).
3. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, *Appl. Spectrosc.*, **43**: 772 (1989).
4. D. E. Honigs, G. M. Hieftje, J. L. Mark, and T. B. Hirschfeld, *Anal. Chem.*, **57**: 2299 (1985).
5. J. Shenk, Equation Selection, in *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality*, G. C. Marten, F. E. Barton II, and J. S. Shenk (eds.), USDA Ag. Hdbk. No. 643, National Technical Information Service, 5285 Port Royal Road, Springfield, VA, 1986, pp. 26–27.
6. W. J. Youden, *Statistical Manual of the AOAC*. AOAC, 1111 N. Nineteenth St., Suite 210, Arlington, VA 22209.
7. H. Mark and J. Workman, Jr., *Anal. Chem.*, **58**: 1454 (1986).
8. N. R. Draper and A. Smith, *Applied Regression Analysis*, John Wiley and Sons, New York, 1981, p. 152.
9. I. A. Cowe and J. W. McNicol, *Appl. Spectrosc.*, **39**: 257 (1985).
10. J. Sustek, *Anal. Chem.*, **46**: 1676 (1974).
11. D. E. Honigs, J. M. Freelin, G. M. Hieftje, and T. B. Hirschfeld, *Appl. Spectrosc.*, **37**: 491 (1983).
12. M. A. Maris, C. W. Brown, and D. S. Lavery, *Anal. Chem.*, **55**: 1694 (1983).
13. H. J. Kisner, C. W. Brown, and G. J. Kavarnos, *Anal. Chem.*, **55**: 1703 (1983).
14. M. Otto and W. Wegscheider, *Anal. Chem.*, **57**: 63 (1985).
15. S. D. Frans and J. M. Harris, *Anal. Chem.*, **57**: 2680 (1985).
16. D. Metzler, C. M. Harris, R. L. Reeves, W. H. Lawton, and M. S. Maggio, *Anal. Chem.*, **49**: 864A (1977).
17. R. N. Cochran and F. H. Horne, *Anal. Chem.*, **49**: 846 (1977).
18. I. A. Cowe and J. W. McNicol, *Appl. Spectrosc.*, **39**: 257 (1985).
19. S. Kawata, H. Komeda, K. Sasaki, and S. Minami, *Appl. Spectrosc.*, **39**: 610 (1985).
20. H. Mark, *Anal. Chem.*, **58**: 2814 (1986).
21. S. Wold, P. Geladi, K. Esbensen, and J. Ohman, *J. Chemometrics*, **1**: 41 (1987).
22. P. Geladi and B. Kowalski, *Anal. Chem. Acta.*, **185**: 19 (1986).
23. D. Haaland and E. Thomas, *Anal. Chem.*, **60**: 1193 (1988).
24. D. Haaland and E. Thomas, *Anal. Chem.*, **60**: 1202 (1988).
25. W. Lindberg, J.-A. Persson, and S. Wold, *Anal. Chem.*, **55**: 643 (1983).
26. I. E. Frank, J. H. Kavilas, and B. R. Kowalski, *Anal. Chem.*, **55**: 1800 (1983).
27. A. Lorber, L. E. Wangen, and B. R. Kowalski, *J. Chemometrics*, **1**: 19 (1987).
28. H. Martens and T. Naes, Multivariate Calibration by Data Compression, in *Near-Infrared Technology in the Agricultural and Food Industries*, P. Williams and K. Norris (eds.), Am. Assoc. Cereal Chemists, St. Paul, MN, 1987, pp. 57–87.
29. ASTM E1655-00, “Standard Practices for Infrared, Multivariate, Quantitative Analyses,” *American Society for Testing and Materials*, 100 Barr Harbor Dr., West Conshohocken, PA 19428.

30. J. Workman and J. Brown, "A New Standard Practice for Multivariate, Quantitative Infrared Analysis, Part 1," *Spectroscopy*, 11: 48–51 (1996).
31. J. Workman and J. Brown, "A New Standard Practice for Multivariate, Quantitative Infrared Analysis, Part 2," *Spectroscopy*, 11: 24–30 (1996).
32. J. Workman, "The First ASTM Standard Practice for NIR Multivariate Quantitative Analysis," *NIR News*, 9: 5–6 (1998).
33. J. Shenk, personal communication, 1986.

8 Data Analysis: Multilinear Regression and Principal Component Analysis

Howard Mark

CONTENTS

8.1	Principles of Calibration: The Error-Free Case	152
8.2	Calibration with Error	156
8.3	Calibration with Error in Y Only	156
8.4	Auxiliary Statistics from Regression	159
8.4.1	Standard Error of Estimate	160
8.4.2	Multiple Correlation Coefficient	160
8.4.3	F for Regression	160
8.4.4	Student's t for the Coefficients	161
8.5	Effect of Intercorrelation	161
8.6	Wavelength Selection in the Single-Wavelength Case	163
8.7	Wavelength Selection with Multiple Wavelengths	163
8.8	Wavelength Selection with Error in Y Only	163
8.9	Wavelength Selection with Noise in the Optical Data	164
8.10	Including Extraneous Variables in a Calibration	166
8.11	Principal Components: Introduction	166
8.11.1	Definition and Discussion	167
8.11.2	Comparison with Fourier Analysis	168
8.11.2.1	Fourier Analysis	168
8.11.2.2	Principal Component Analysis	171
8.11.3	What Makes Principal Components Unique?	176
8.11.4	Are Principal Components Brought by the Stork?	178
8.12	Calibration	178
8.13	Characteristics	182
8.14	Spectral Reconstruction	182
8.15	Advantages	184
8.15.1	Lack of Requirement for Wavelength Selection	184
8.15.2	Robustness	185
8.15.3	Interpretation of Calibration Equations	185
8.16	Qualitative Analysis (SIMCA)	185
8.17	Data Compression	185

8.18 What Principal Components Won't Do	186
Acknowledgment	187
References	187

8.1 PRINCIPLES OF CALIBRATION: THE ERROR-FREE CASE

We will begin by considering a simplified model, one where Beer's law strictly applies, the constituent of interest (which is the only one in the sample) has only a single absorption band and is contained in a nonabsorbing matrix, the concentration of the constituent is known exactly over a fairly broad range, and the spectrometer we are using has absolutely no noise, nonlinearity, or any other fault. In such an ideal case the spectrum would look like that in Figure 8.1a.

In this idealized case, the height of the absorbance peak is strictly proportional to the concentration of the constituent. A plot of these variables would look like Figure 8.2. In a case like the one portrayed in Figure 8.1a and Figure 8.2, where all the assumptions hold exactly, we can calibrate our system using only two samples since two points determine the line. Then the slope and intercept can be easily determined from the usual mathematical formulas, and then we can write the usual formula for Beer's law

$$A = \varepsilon C \quad (8.1)$$

where ε represents the molar extinction coefficient (we have assumed unity pathlength in order to simplify the equation) and is thus equal to m , the slope of the line.

However, in the usual case it is not the absorbance we are interested in, but rather the concentration of the constituent. We make this explicit by solving Equation (8.1) for concentration

$$C = \varepsilon^{-1} A \quad (8.2)$$

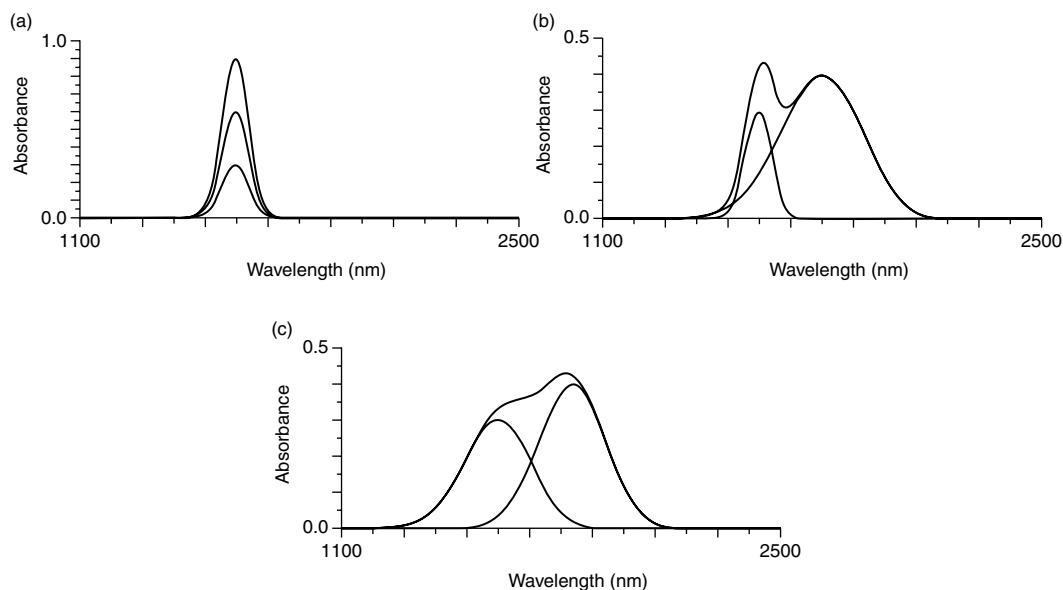


FIGURE 8.1 Idealized spectra, with no noise. (a) Three different concentrations of a material with only one absorbance band in a nonabsorbing solvent. (b) Analyte is similar to (a), with an interfering material also present. (c) Bands of both materials interfere with each other.

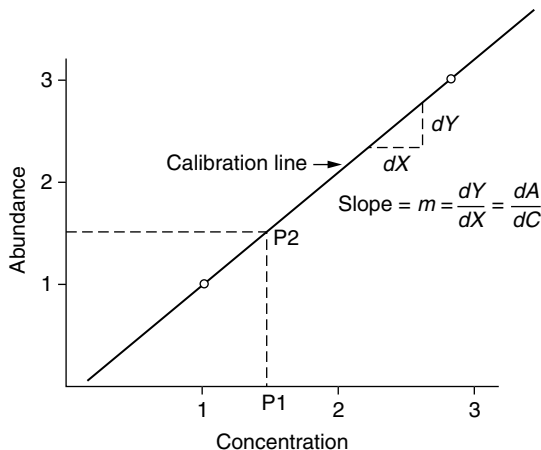


FIGURE 8.2 In the strict Beer's law case, with no error in either variable, a straight line can be plotted exactly through all the data, since two points determine the line. An unknown can then be measured by measuring its absorbance at P1, projecting the reading to the line at point P2, and projecting the point of intersection to the concentration axis.

where ε^{-1} is the inverse of ε . This expression is sometimes called the “inverse Beer's law” formulation of the situation. We note here that ε^{-1} is a constant and thus could be replaced by any symbol representing that constant. We will find it convenient to use the symbol b . Thus Equation (8.2) can be rewritten as

$$C = bA \quad (8.3)$$

where b represents the constant calculated from the inverse of ε .

The next simplest case is illustrated in Figure 8.1b. In this case the absorption band of the constituent that we wish to measure is overlapped by a band of an interferent. Since the absorption at λ_2 is due only to constituent two, we could, if we wished, compute the concentration of this material as we did previously. The total absorption at λ_1 , however, is due to the sum of the contributions of both constituents at that wavelength. If we knew the absorbance due to constituent one only we could use Equation (8.3) to calculate the concentration of constituent one. We can find the value of this absorbance by subtracting the absorbance of constituent two at wavelength one

$$C_1 = \frac{1}{\varepsilon_{11}}(A_{11} - C_2\varepsilon_{12}) \quad (8.4)$$

where the double (i, j)th subscript on ε represents the extinction coefficient at λ_i of constituent j .

However, C_2 is unknown. Yet by measuring at λ_2 we can obtain a measure of C_2 :

$$A_{22} = \varepsilon_{22}C_2 \quad (8.5)$$

Substituting Equation (8.5) into Equation (8.4) we obtain

$$C_1 = \frac{1}{\varepsilon_{11}}A_{11} - \frac{\varepsilon_{12}}{\varepsilon_{11}\varepsilon_{22}}A_{22} \quad (8.6)$$

We have explicitly solved for the concentration of C_1 in terms of only the measured absorbances at λ_1 and λ_2 and the extinction coefficients of the two materials at the two measured wavelengths. We note that this is also an inverse Beer's law formulation, and that at no time do we ever have to know the concentration of constituent two explicitly. Thus we find that the inverse Beer's law

formulation of quantitative spectroscopic analysis has several distinct advantages over the direct Beer's law formulation:

1. The equations developed provide an explicit solution for the concentration of the unknown analyte in terms solely of the measurable absorbances.
2. Only the concentrations of the desired analyte need be known. The values for other constituents of the sample need never be determined, even during the calibration step, although in order to develop a good calibration equation there should be variation in these values. Indeed, in some cases, even the number of other compounds in the calibration samples may not be known.
3. It is unnecessary to have values for the various extinction coefficients.
4. In some types of measurements, and diffuse reflectance provides a prototype example, the effects of extraneous physical phenomena are super-imposed on the readings. In such cases distinct values for extinction coefficients cannot be specified, but the inverse Beer's law formulation can account for these phenomena by implicitly including corrections for the effects in the calibration coefficients.

When we consider the effect of error in the data we will find that there is another important, although more subtle, advantage of applying the inverse Beer's law formulation to quantitative spectroscopic analysis.

The next case is illustrated in Figure 8.1c, where we have two materials in a sample that absorb at mutually interfering wavelengths. In this case Beer's law still holds, and we still assume no error of any sort in either the optical data or the constituent values. Since Beer's law holds, the spectrum we can measure is, at every wavelength, the sum of the absorbances of the spectra of the constituents. Thus, at wavelengths λ_1 and λ_2 , we can write the Beer's law expression for the corresponding absorbances:

$$A_1 = \varepsilon_{11}C_1 + \varepsilon_{12}C_2 \quad (8.7a)$$

$$A_2 = \varepsilon_{21}C_1 + \varepsilon_{22}C_2 \quad (8.7b)$$

In Equation (8.7a) and Equation (8.7b) the subscripts on the absorbances (A) represent the absorbance at the two different wavelengths, the subscripts on the constituents (C) represent the two different constituents, and the i, j subscripts represent the molar extinction coefficients for the two materials (the j subscript) at the two different wavelengths (the i subscript).

Is it possible, for the cases of Figure 8.1b and Figure 8.1c, to draw a graph that looks like Figure 8.2? Conceptually the answer is yes, although it is extremely difficult to draw a graph that could actually be used to read values from the way we could use Figure 8.2.

Since there are two optical variables and a concentration variable, there are a total of three variables. This system thus requires a three-dimensional space for representation. Such a space is representable on a two-dimensional piece of paper. When more complicated situations are encountered, higher dimensional spaces become necessary; these can no longer be represented with two-dimensional diagrams. However, the mathematical basis for handling these cases is the same as for the two wavelength (three-dimensional) case; thus this case serves as an ideal intermediate example. Such a graph is shown in Figure 8.3.

As we see in the figure, three points are sufficient to determine a plane; this plane is the calibration plane, completely analogous to the calibration line that we presented for the spectrum of Figure 8.1a. Its use is also analogous: to measure an unknown, the absorbances at the two analytical wavelengths are measured and the point P1 determined. Just as in the one-wavelength case, P1 is projected to the calibration plane, and the point of intersection (P2) is projected onto the concentration axis, thus providing the reading.

How can we do this mathematically, in a manner similar to the conversion we made in going to Equation (8.3)? Well, Equation (8.7a) and Equation (8.7b) are two equations in two unknowns,

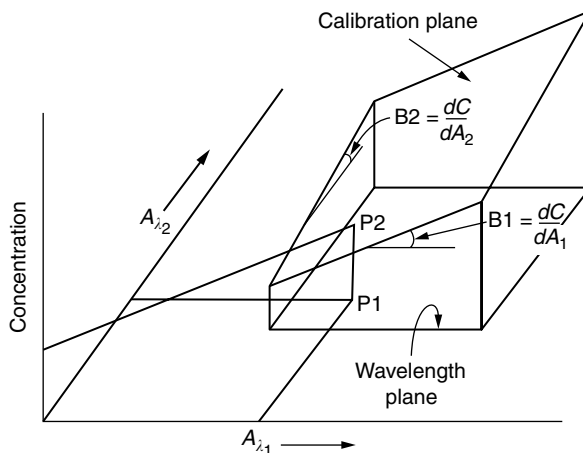


FIGURE 8.3 When interferences require the use of more than one wavelength, the graphical approach requires a *calibration plane* instead of a *calibration line*.

the unknowns in this case being the concentrations. We can solve these equations for the two concentrations C_1 and C_2 . The result is

$$C_1 = \left(\frac{\varepsilon_{22}}{\varepsilon_{11}\varepsilon_{22} - \varepsilon_{12}\varepsilon_{21}} \right) A_1 + \left(\frac{-\varepsilon_{12}}{\varepsilon_{11}\varepsilon_{22} - \varepsilon_{12}\varepsilon_{21}} \right) A_2 \quad (8.8a)$$

$$C_2 = \left(\frac{-\varepsilon_{21}}{\varepsilon_{11}\varepsilon_{22} - \varepsilon_{12}\varepsilon_{21}} \right) A_1 + \left(\frac{\varepsilon_{11}}{\varepsilon_{11}\varepsilon_{22} - \varepsilon_{12}\varepsilon_{21}} \right) A_2 \quad (8.8b)$$

Equation (8.8a) and Equation (8.8b) display many characteristics that are important in understanding what is going on because they represent the prototype equations for multiwavelength analysis.

First of all, we note again that the concentrations of the constituents are expressed solely in terms of the measurable quantities A_1 and A_2 , and that they are related to these measurable quantities by factors (i.e., the coefficients of the absorbances) that are combinations of constants. Therefore these factors are themselves constant, and thus Equation (8.8a) and Equation (8.8b) can be rewritten as

$$C_1 = b_{11}A_1 + b_{12}A_2 \quad (8.9a)$$

$$C_2 = b_{21}A_1 + b_{22}A_2 \quad (8.9b)$$

Basically, what we have done is to solve the simultaneous Equation (8.7a) and Equation (8.7b) for the unknowns, which are the concentrations. In principle we could do the same thing for any number of wavelengths needed to perform a spectroscopic analysis. The difficulty with that approach is that the equations quickly become far too cumbersome to manage. Three is the practical limit here; this approach is useful mainly for gaining understanding.

To extend our capabilities beyond two or three wavelengths, we take note of the details of the conversion from Beer's law to inverse Beer's law: the form of the coefficients of the absorbances in Equation (8.8a) and Equation (8.8b). These coefficients are the result of applying the process known as *matrix inversion* to a 2×2 array of numbers. Matrix operations are beyond the scope of this chapter, so we will simply note that they are the means of generalizing these concepts to larger numbers of wavelengths.

One further point merits discussion. In Equation (8.8a) and Equation (8.8b) we have solved Equation (8.7a) and Equation (8.7b) for both constituents. This is useful, but not required. As in the

case of Figure 8.1b, if we are interested in the concentration of only one of the constituents we are not required to solve for the other; the concentration of the second constituent need never be known.

8.2 CALIBRATION WITH ERROR

Errors in the data to be used for calibration may be classified into three categories: random error in the reference laboratory values, random error in the optical data, and systematic error in the relationship between the two. The proper method of dealing with errors of the data depends on whether the affected variables are the reference values or the optical data.

Theoretically, if all the necessary extinction coefficients are known, then error in the data would have no bearing on the generation of a suitable calibration; Equation (8.8) could be created from first principles and used to calculate the calibration coefficients needed to predict unknown samples via Equation (8.9). In reality, knowledge of all these factors has not yet been achieved, particularly in the near-infrared (NIR) region of the spectrum. Thus calibrations are performed empirically; in this situation, the question of error in the data becomes important.

In fact, it is exceedingly difficult to decide which error sources are larger: the optical or reference laboratory's. The difficulty comes about because of the different units in which the two are measured; how is one to compare error in percent constituent with electronic noise level to see which is larger?

However, the noise of current NIR instrumentation is so extraordinarily small that it is invariably smaller than almost anything else in the calibration experiment. That fact might be considered justification for concluding a priori that the reference laboratory results are always the dominating factor. However, that simplistic conclusion ignores the fact that the total error of the optical data depends on sample-induced errors as well as instrumental errors, and these can be much larger than the noise level of the instrument. Such errors include particle size effects, repack variations, effects of impurities, and effects due to changing physical characteristics of the samples (e.g., crystallinity).

The knowledge that in a large number of cases the errors in the optical data due to the samples are also small has made routine the implicit assumption that the error in the optical data is always small compared to the reference laboratory errors, so that the inverse Beer's law formulation is always used; thus the optical data are always considered the *X* data and the reference laboratory values are considered the *Y* data.

8.3 CALIBRATION WITH ERROR IN *Y* ONLY

In the simplest and most studied case there is error in only the *Y* (the dependent) variable (i.e., in the reference laboratory values for the constituent). In spectroscopic practice this situation frequently arises when natural or synthetic products must be analyzed by a reference method in order to obtain values for the calibration, or training, set of samples. Figure 8.4a illustrates the danger of using only two points to determine the calibration line in the presence of error. As calibration line 1 shows, it is very unlikely that the points chosen will actually correspond to the proper slope and intercept of the line that the majority of points follow. If the points chosen are particularly inopportune, a line such as shown by calibration line 2 might even be obtained. In a case such as this, the desired calibration line is one such as illustrated in Figure 8.4b, where a "best fitting" calibration line is determined in such a way that all the data are taken into account. In practice, the best fitting line is determined by the method of least-squares regression.

The multiwavelength situation corresponding to Figure 8.4b is illustrated in Figure 8.5. Just as the line in Figure 8.4b was fitted to all the data for the single wavelength, the plane in Figure 8.5 is fitted to all the data in the multiwavelength case, splitting the differences among them.

The standard method of calibration under these conditions is the use of least-squares regression analysis. Regression is particularly suitable when the reference values for the analyte are determined by chemical analysis using a reference method. Regression then becomes the preferred method for a number of reasons: only the constituents for which calibrations are desired need be determined by

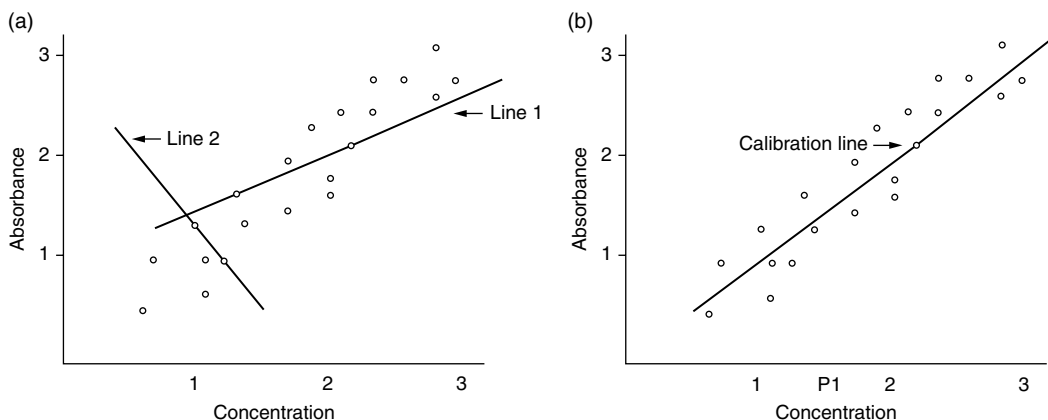


FIGURE 8.4 (a) When there is error in the data, using only two points to determine the line can result in a poorly fitting line as these two lines illustrate. (b) By getting a best fit to all the points, much better results will be obtained.

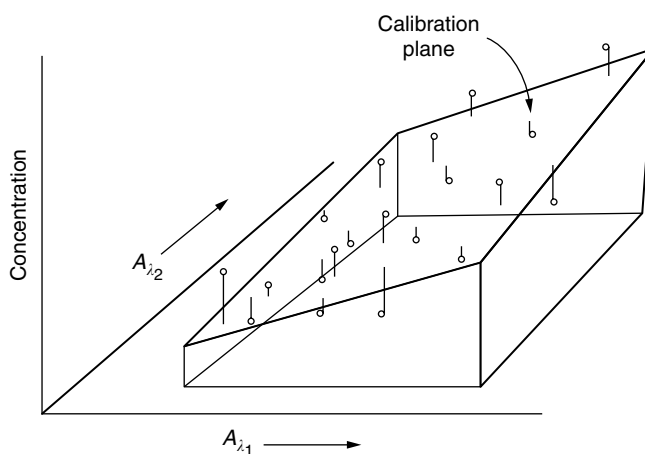


FIGURE 8.5 When an error is present in the data, the calibration plane must be fit by least squares.

the reference method; since the reference values usually contain the largest errors, making these the Y variables puts the errors in the proper place in the data, and, while simple regression is performed on data with symmetry between the two variables, the data used for multiple regression do not have the same symmetry under these circumstances. For example, in Figure 8.2, the two variables are equivalent and, as Equation (8.7) and Equation (8.8) show, can be related by equivalent mathematical formulas. On the other hand, if reference values for only one or a few constituents are available for the samples, it is not possible to write a set of equations corresponding to Equation (8.7) for the multivariate case; the symmetry of the two-variable system has been broken.

Regression analysis, both simple regression (meaning only one independent variable) and multiple regression (more than one independent variable), has been developed over many years, and many of its characteristics are known. In particular, the least-squares method is valid only when the following four fundamental assumptions hold:

1. There is no error in the independent (X) variables.
2. The error in the dependent variable (Y) is normally distributed.

3. The error in the dependent variable is constant for all values of the variable.
4. The system of equations is linear in the coefficients (*Note: not necessarily in the data; the variables representing the data may be of polynomial or even transcendental form*).

Here we present the derivation of the equations for multiple regression analysis (the “least-squares” fit of a multivariate equation to data). We note that the derivation is subject to these four assumptions being valid. Methods of testing these assumptions and advice on dealing with failure of them to apply in particular cases are well discussed in the statistical literature. The reader is strongly urged to become familiar with methods of inspecting data and to apply them routinely when performing regression calculations.

A more detailed derivation, including the derivation of the auxiliary statistics for a regression, can be found in *Applied Regression Analysis*, by N. Draper and H. Smith (John Wiley and Sons, 2nd ed., 1981), a book that should be considered required reading for anyone with more than the most casual interest in regression analysis.

A dataset to be used for calibration via the multiple regression least-squares method contains data for some number (n) of readings, each reading presumably corresponding to a specimen, and some number (m) of independent variables, corresponding to the optical data. The dataset also contains the values for the dependent variable: the analyte values from the reference laboratory. We begin by defining the error in an analysis as the difference between the reference laboratory value of the analyte, which we call Y , and the instrumental value for the constituent, which we call \hat{Y} (read “ Y -hat”):

$$e = Y - \hat{Y} \quad (8.10)$$

Using X to designate the values of the optical data, we note that, for any given sample, \hat{Y} is calculated according to the calibration equation

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \cdots + b_mX_m \quad (8.11)$$

and thus the error e is equal to:

$$e = Y - b_0 - b_1X_1 - b_2X_2 - \cdots \quad (8.12)$$

Equation (8.12) gives the error for a single reading of a single specimen in the set of specimens composing the calibration set. The least-squares principle indicates that we wish to minimize the sum of the squares of the errors for the whole set. The sum of the squares of the errors is

$$\sum_{i=1}^n e^2 = \sum_{i=1}^n (Y - b_0 - b_1X_1 - b_2X_2 - \cdots - b_mX_m)^2 \quad (8.13)$$

where the subscripts on the variables represent the different wavelengths, and the summations are taken over all the specimens in the set.

In order to minimize the sum of the squares of the errors, we do the usual mathematical exercise of taking the derivative and setting it equal to zero. For a given set of data, the error, and the sum squared error, will vary as the coefficients change, therefore the derivatives are taken with respect to each coefficient.

$$\frac{\partial(\sum e^2)}{\partial b_0} = \frac{\partial}{\partial b_0} \sum (Y - b_0 - b_1X_1 - b_2X_2 - \cdots)^2 = 0 \quad (8.14a)$$

$$\frac{\partial(\sum e^2)}{\partial b_1} = \frac{\partial}{\partial b_1} \sum (Y - b_0 - b_1X_1 - b_2X_2 - \cdots)^2 = 0 \quad (8.14b)$$

$$\frac{\partial(\sum e^2)}{\partial b_2} = \frac{\partial}{\partial b_2} \sum (Y - b_0 - b_1X_1 - b_2X_2 - \cdots)^2 = 0 \quad (8.14c)$$

Taking the specified derivatives gives rise to the set of equations:

$$\sum 2(Y - b_0 - b_1X_1 - b_2X_2 - \dots) = 0 \quad (8.15a)$$

$$\sum 2X_1(Y - b_0 - b_1X_1 - b_2X_2 - \dots) = 0 \quad (8.15b)$$

$$\sum 2X_2(Y - b_0 - b_1X_1 - b_2X_2 - \dots) = 0 \quad (8.15c)$$

The next step is to divide both sides of Equation (8.15) by two, and simplify by multiplying out the expressions:

$$\sum (Y - b_0 - b_1X_1 - b_2X_2 - \dots) = 0 \quad (8.16a)$$

$$\sum (YX_1 - b_0X_1 - b_1X_1X_1 - b_2X_1X_2 - \dots) = 0 \quad (8.16b)$$

$$\sum (YX_2 - b_0X_2 - b_1X_2X_1 - b_2X_2X_2 - \dots) = 0 \quad (8.16c)$$

The next step is to separate the summations and rearrange the equations:

$$\sum b_0 + b_1 \sum X_1 + b_2 \sum X_2 + \dots = \sum Y \quad (8.17a)$$

$$b_0 \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1X_2 + \dots = \sum X_1Y \quad (8.17b)$$

$$b_0 \sum X_2 + b_1 \sum X_1X_2 + b_2 \sum X_2^2 + \dots = \sum X_2Y \quad (8.17c)$$

The multipliers of the b_i on the left-hand side of Equation (8.17), as well as the terms of the right-hand side of Equation (8.17), are all constants that can be calculated from the measured data. Thus we have arrived at a system of $m + 1$ equations in $m + 1$ unknowns, the unknowns being the b_i , and the coefficients of the b_i being the summations. Solving these equations thus gives us the values for the b_i that minimizes the sum-squared errors of the original set of equations represented by Equation (8.13).

It is important to remind ourselves again at this point that the error of the data indicated in Figure 8.4b and Figure 8.5 is only in the Y (dependent) variable. In the vast majority of cases we deal with in near-infrared analysis (NIRA), this situation is found to obtain. Regression theory states that the standard error of estimate (SEE), the measure of error in the regression, should be equal to the error of the dependent variable. Indeed, it is because there is so little error in the instrumental measurements on NIRA that multiple regression analysis is an appropriate mathematical tool for calibrating the instruments.

8.4 AUXILIARY STATISTICS FROM REGRESSION

In addition to the calibration coefficients, most multiple regression programs calculate a number of auxiliary statistics, which have the purpose of helping the operator decide how well the calibration fits the data, and how well the calibration can be expected to predict future samples. The most important of these statistics are the SEE, the multiple correlation coefficient (R), the statistic F for regression, and the Student t values for the regression coefficients.

8.4.1 STANDARD ERROR OF ESTIMATE

The SEE and the multiple correlation coefficient are both ways of calculating how well the calibration equation fits the data. The key difference between them is that the SEE has units, while the multiple correlation coefficient is dimensionless. The units of the SEE are the same as the units of the dependent variable.

The SEE is calculated according to the equation:

$$SEE = \sqrt{\frac{\sum_{i=1}^n (Y - \hat{Y})^2}{(n - m - 1)}} \quad (8.18)$$

where Y is the reference laboratory value of the analyte, \hat{Y} is the analyte value produced by the calibration equation, n is the number of readings in the data, and m is the number of independent variables in the calibration equation. This equation calculates what is sometimes called the RMS error, corrected for what are known as degrees of freedom. This correction takes into account the fact that when many variables are used for fitting the data, the fit to the data can appear to be much better than it actually is. The utility of this formulation is that it depends on the differences between the instrumental value for the analyte and the reference laboratory value. In the absence of instrumental error, the SEE is a measure of the error of the reference laboratory. It is also an indication of whether the answers provided by the calibration equation will be sufficiently accurate for the purposes for which they are being generated.

It is possible for the SEE to indicate that the calibration is less accurate than the reference laboratory, even in the absence of instrumental error, if the wavelengths used as the independent variables in the calibration do not account for all the interferences in the samples or if extraneous physical phenomena are present.

8.4.2 MULTIPLE CORRELATION COEFFICIENT

The multiple correlation coefficient is, as previously stated, a dimensionless measure of the degree to which the calibration fits the data. The multiple correlation coefficient is the same as the ordinary correlation coefficient between the reference laboratory values and the instrument readings for the samples in the calibration set. Normally a correlation coefficient can have values between -1 and $+1$. In a calibration situation, however, only positive values can exist. A value of this statistic close to zero indicates that the calibration is failing to relate the instrument readings to the reference values. As the correlation coefficient increases, the instrument readings become better and better indicators of the reference values until, when it reaches unity, the instrument values and the reference values are identical in all cases.

Because the multiple correlation coefficient is dimensionless, it is a useful way of comparing data or results that have different units, and that are difficult to compare in other ways. On the other hand, the multiple correlation coefficient is a number, and just looking at a number doesn't necessarily tell all about a situation. In particular, the value of the multiple correlation coefficient does not give any indication of how reliable that number is or how well the calibration equation can be expected to perform on future samples.

8.4.3 F FOR REGRESSION

Since the multiple correlation coefficient alone does not indicate how reliable the data are, it is useful to have a statistic that performs that function. Such an indicator has been developed by statisticians: it is the F for regression. F is calculated from the expression:

$$F = \frac{[\sum (\hat{Y} - \bar{Y})^2]/m}{[\sum (Y - \hat{Y})^2]/(n - m - 1)} \quad (8.19)$$

Note that the summation in the denominator is the same one that appeared in the expression for SEE. In terms of giving a number that can be related to the multiple correlation coefficient, the F for regression can also be shown to be equal to

$$F = \frac{R^2}{(1 - R)^2} \times \frac{n - m - 1}{m} \quad (8.20)$$

From Equation (8.20) it is clear that this statistic becomes large when R is close to one (there is a close relationship between the instrumental values and the reference laboratory values), if n is large (many specimens are used in the calibration) and m is small (few independent variables are needed for the calibration, avoiding the possibility of equations that appear good but are actually overfitting the data).

What do values of F for regression mean in terms of the analytical performance of the calibration? We have noted previously that the SEE measures the accuracy of the calibration equation for the calibration samples. The value of F is an indicator of how well this accuracy can be expected to hold up when unknown samples, not included in the calibration, are measured. If the value of F for regression is low, the calibration is not robust, and the performance on future unknown samples is likely to be substantially worse than on the calibration samples. If F is large, then the calibration should have approximately the same accuracy on future unknown samples as on the calibration samples.

8.4.4 STUDENT'S t FOR THE COEFFICIENTS

The value of t for a given coefficient is equal to the value of the coefficient divided by the estimate of the standard error of that coefficient obtained from the regression data. It arises this way: because the values of the dependent variable each contain a contribution from the random error in the reference laboratory, if another set of samples were to be used for the calibration, a different set of data would be used in the calibration, even if the data were taken from different aliquots of the same specimens. Then, when the calculations were performed, slightly different answers would result. It is possible to estimate from the data itself how much variation in the coefficients would result from this source. The estimate of the variation is called the “standard error of the coefficient;” and Student's t is a coefficient divided by its standard error.

Thus, if Student's t is large, then the potential variation is small, and vice versa. While there are many possible reasons why t for a particular coefficient might be small, a large value for t indicates several things: the corresponding coefficient is relatively unaffected by variations in the data; the coefficient is well defined and thus likely to maintain its value if measurement conditions change somewhat, making the calibration equation robust against small changes in the measurement conditions. A large value of t also indicates that the corresponding independent variable is important to the overall predictive capability of the calibration equation. It is also useful to note that this t is the square root of a statistic known as the partial F for deletion, another statistic sometimes used to determine whether a coefficient is a candidate for deletion from a calibration equation.

8.5 EFFECT OF INTERCORRELATION

Even in the absence of error in the X variables, intercorrelation among them can occur, with deleterious effects on the calibration. Intercorrelation among the independent variables (sometimes called *multicollinearity*) is often found in data taken using diffuse reflectance techniques, due to what is commonly referred to as *particle size effects*, and contributes to the instability of calibrations performed on such data. When even one of the calibration wavelengths changes, the calibration coefficients at all wavelengths change due to the variation in extinction coefficients. In addition to that source of variation, intercorrelation among the optical data used as independent variables in a

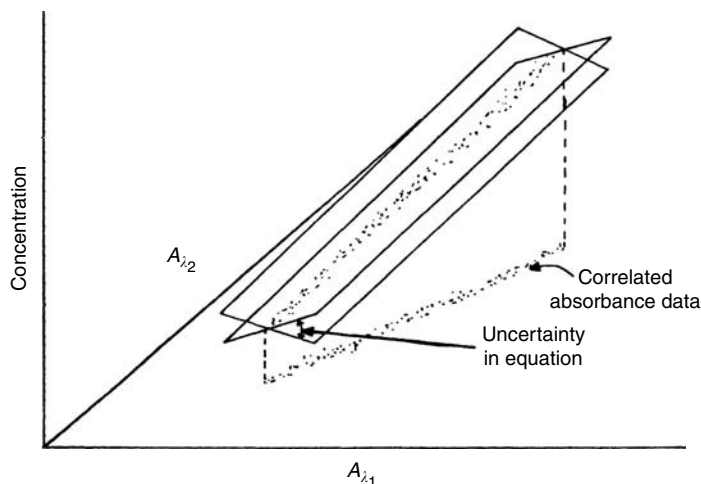


FIGURE 8.6 Correlation among the optical data causes the calibration plane to be unstable. This results in large apparent differences between calibrations resulting from different data sets.

calibration makes the calibration coefficients sensitive to the values of the dependent variable against which the instrument is being calibrated.

The source of this sensitivity can be seen in Figure 8.6, where the condition is displayed diagrammatically. We previously alluded to the necessity of having the data corresponding to the independent variables spread out over the plane of the variables; Figure 8.5 showed such an ideal situation. Figure 8.6 shows a two-wavelength (three variable) situation where the optical data at the two wavelengths are shown highly correlated with each other rather than being spread out uniformly over the entire λ_1 to λ_2 plane. Thus the reference laboratory data for the constituent, instead of being spread out over the calibration plane, is concentrated into a narrow, somewhat cigar-shaped region of space. The axis of this cigar-shaped region of space is well defined, but directions orthogonal to it are not. Consequently, when a calibration plane is calculated, it is very unstable around the axis of the cigar shape. This is manifested by the fact that when different sets of data are used to perform a calibration, small differences in the random errors of the reference laboratory tilt the plane severely, causing vastly different calibration planes to be calculated. This shows up in the calibrations as widely differing sets of calibration coefficients. The data analyst, seeing only the changes in the calibration coefficients, cannot be sure whether these changes are due to the intercorrelation phenomena described here, real differences in the samples being used for the calibration, or other artifacts.

Figure 8.6 shows the situation for a two-wavelength calibration. Intercorrelation effects can also occur when more than two wavelengths are being used. Indeed, as more wavelengths are included in the calibration, this effect is likely to become more common and more severe. In Figure 8.6, the degree of intercorrelation can be described mathematically by the simple correlation coefficient between the data at the two wavelengths. When more than two wavelengths are included in the calibration, the corresponding mathematical description is the multiple correlation coefficient among the data at all the wavelengths involved. Conceptually, this can be generated by using the data at one of the wavelengths as the dependent variable in a multiple regression calculation, with the data at all the remaining wavelengths used as the independent variables. The multiple regression algorithm applied to such data can produce all the auxiliary statistics for the regression; of these, only the multiple correlation coefficient is of interest here. It is that statistic that determines how unstable the calibration coefficients are. If used, this calculation should be applied to the data at each of the wavelengths in the calibration; the resulting multiple correlation coefficient for each wavelength will indicate the degree of instability of the calibration coefficient corresponding to the data at that wavelength.

8.6 WAVELENGTH SELECTION IN THE SINGLE-WAVELENGTH CASE

At this point it is appropriate to consider the question of wavelength selection. We start with the noise-free case: what wavelength should we choose for the analysis? Intuitively, most spectroscopists would select the peak of the absorbance band in Figure 8.1a. If pressed for a reason, the one given would usually be something related to “best signal-to-noise.” But in the hypothetical, ideal case we are considering, *there is no noise*. Hence the signal-to-noise ratio is constant all across the absorbance band, thus providing no reason to prefer one wavelength over another within this spectral range.

However, neither does that mean that we can choose a wavelength completely arbitrarily. When performing the calibration, the value of the calibration coefficient will depend on the value of ϵ at the analytical wavelength, by Equation (8.2). Thus the wavelength selected for calibration must be used for analysis or, since the analyte will have a different extinction coefficient at a different wavelength, error will be generated by lack of fit of the equation due to the use of an incorrect value of the calibration coefficient.

8.7 WAVELENGTH SELECTION WITH MULTIPLE WAVELENGTHS

Even with multiple wavelengths, in the perfect error-free case, wavelength selection for the purpose of optimizing calibration/prediction performance is again unnecessary and/or impossible. For the illustrative examples, we have shown the wavelengths used as being located at the absorption maxima of the constituent bands, but this was a matter of convenience, habit, and utility. If the data are truly perfectly error-free, then any wavelengths will permit the generation of a calibration equation that fits the data perfectly; two points determine the calibration line and three, the plane with no error regardless of the wavelength used. Thus wavelength selection based on analytical performance is impossible because that provides no criterion for choosing one wavelength over another, as long as we remain in the wavelength regions where the constituents do in fact exhibit absorption. Wavelength selection is unnecessary for the same reason, in other words, any wavelength will do the job equally well.

Different wavelengths will, of course, give rise to different values for the calibration coefficients, reflecting the different values of the extinction coefficients at different wavelengths. From Equation (8.8a) and Equation (8.8b) it is clear that changing one wavelength will result in a change in the coefficients for all wavelengths, since the calculation for each calibration coefficient contains a contribution from each constituent at each wavelength.

In the case shown in Figure 8.1c, with mutually interfering peaks, we have even lost the justification we previously had for using the spectral peaks. As Figure 8.1c shows, with mutually interfering peaks a wavelength corresponding to an absorption peak of one material will inevitably correspond to the slope of the other material. The change in extinction coefficient on the slope of the second material will inevitably change all the calibration coefficients.

8.8 WAVELENGTH SELECTION WITH ERROR IN Y ONLY

The situation here is similar to the completely error-free case discussed previously. Even with error in the Y variable, if the X variables are noise-free and the relationship between the X 's and the Y is linear, there is no criterion for choosing one wavelength over another as the predictor variable.

The mathematical operation called analysis of variance (ANOVA) provides a mathematical proof that the total variance of the error of a result is the sum of the variances due to each source of error.

Thus, if the errors in X are in fact zero, then the signal-to-noise ratio is the same at every wavelength, although the signal is not what we might think it is. This arises from the fact that the error in reference laboratory values are of course constant regardless of the wavelengths used for the optical data, and since (by hypothesis) there is no error contribution from the optical data there is no change in the signal-to-noise ratio as different wavelengths are used, and thus there is again no criterion for making a selection.

This is the cause of the notorious propensity NIRA has for selecting different wavelength sets when performing calibrations on different datasets, even though these sets contain essentially equivalent data for the same constituent in the same material: without a sharply defined criterion, the computer is essentially “dithering” over a number of different but equivalent possibilities. However, because of the effects of changes in ε due to the different wavelength combinations, the changes in wavelength give rise to calibrations with vastly different sets of coefficients.

8.9 WAVELENGTH SELECTION WITH NOISE IN THE OPTICAL DATA

We have seen that when all the variables are completely noise free there is no preferred wavelength to use for performing the analysis. For example, Figure 8.7 shows such a spectrum, corresponding to Figure 8.1a. It is clear to the spectroscopist that a wavelength near the peak of the absorbance band is preferable to a wavelength in the tail because of its much higher signal-to-noise ratio. But what does this mean exactly? How can we decide where the noise begins and ends? And, more importantly, what constraints are there on a computer, programmed to pick the “best” wavelength or set of wavelengths?

To investigate these questions, we must review the nature of noise in a spectrum. When a spectrum is noisy, the value measured at any wavelength is the sum of the “true” value plus the contribution of the noise source. Indeed, mathematicians have shown that under such conditions the only factor that has any consistency is the *probability* of finding any given value of the noise, and that the standard

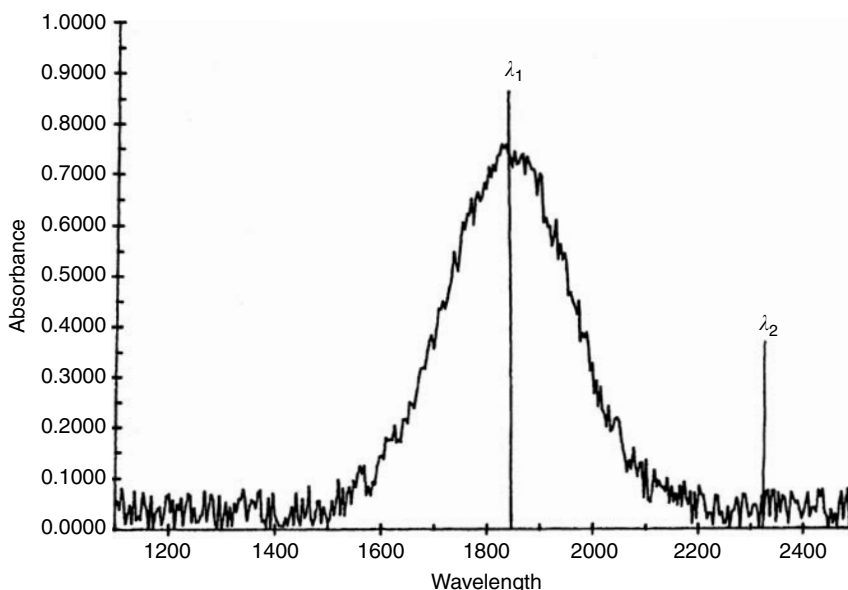


FIGURE 8.7 Noise in the optical data affects the wavelengths that will be chosen by an automatic computer search; a priori predictions of which wavelength will be chosen can be done only on a probabilistic basis.

deviations and variances follow well-defined distributions whose properties have been extensively studied. The distribution followed by variances is called the χ^2 distribution.

The range of values of any statistic that will occur with a known probability is called the *confidence interval* for that statistic. In the case of spectral data contaminated with noise, we need to determine the confidence interval for the wavelength to be chosen.

Consider Figure 8.7. At the peak of the spectral absorption band shown, the spectrum has some signal-to-noise ratio. At the wavelengths near the spectral peak, the “signal” is essentially constant but, as mentioned, the noise varies with wavelength. On the tail of the absorption band, the signal is much smaller than at the peak. Hence, for any given amount of noise, the signal-to-noise ratio will be smaller in the tail than at the peak. The wavelength chosen will be the one where the signal-to-noise ratio is the highest of all the wavelengths within the confines of the absorbance band.

Clearly, when the signal due to the absorbance band falls so low that its ratio to the smallest value of noise at any one wavelength is less than the signal-to-noise ratio at the peak, then no wavelength further away from the peak can be chosen. But the value of noise, and the actual wavelength at which that smallest noise is found, is random and determined by the probabilistic characteristics of the χ^2 distribution of variances.

Another consideration that must be taken into account is the fact that the χ^2 distribution itself is a function of the number of degrees of freedom in the data (degrees of freedom is not exactly the same as the number of readings, but for our current purposes, it is close enough that we will consider them to be the same). The point that needs to be made is that as more readings are included in the data, the lower confidence limit rises and gets closer to the most probable value of the distribution. This translates into the fact that as more readings are collected the confidence interval for wavelength selection gets tighter around the peak wavelength, although this happens very slowly.

This procedure, however, does give us the information needed to put probabilistic limits on the actual wavelengths that will be chosen. For example, from standard statistical tables we find that for 20 df (degrees of freedom), 95% of the time the variance will be above .542 of the expected value and 5% of the time below it. The corresponding ratio of standard deviations is .736. Therefore, 95% of the time the wavelength chosen will fall within the bounds where the absorbance is .736 of the peak absorbance.

The distribution of wavelengths chosen, then, will depend on both the probabilistic distribution of the noise, which depends on the distribution of χ^2 (which in turn depends on the number of readings in the dataset), and on the shape of the absorbance band in question, because that determines the signal part of signal-to-noise.

A computer, of course, knows nothing of this. In the usual procedure for calibrating NIR instrumentation, the computer is given a set of reference laboratory values for a group of specimens and the set of optical data measured from those same specimens, and told to use some variation of an algorithm that instructs it to find the “best” (meaning the smallest SEE) calibration from the available data. The computer then does exactly what it is told, neither knowing nor caring that its choice of wavelength is, to a large extent, based on the noise characteristics of the particular dataset it happens to be working with. This accounts for the notorious propensity of computerized wavelength searches to come up with different sets of wavelengths (usually the “wrong” ones) for what are purported to be the same types of specimens: each one is matched to the particular noise characteristics of the individual dataset that it happens to be using; the wavelengths are effectively being chosen at random.

To overcome this problem a different method of wavelength selection must be used. Two methods that suggest themselves are first, for the operator to override the computer and select wavelengths himself. Current research indicates that when the operator forces the computer to use the same wavelengths on different datasets, the resulting calibrations are essentially the same (i.e., the computer arrives at the same calibration coefficients). The second method is to avoid the use of the concept of a best calibration, since it is a fiction, anyway, and use some other criterion to select the wavelengths.

8.10 INCLUDING EXTRANEOUS VARIABLES IN A CALIBRATION

Most of the specialized packages developed for NIR spectroscopic analysis do make a sharp distinction between the optical data and the constituent (reference laboratory) results. With some programs (particularly older ones), optical data may be used only as independent variables, and constituent values only as dependent variables in a calibration. During the discussion of the effects of intercorrelation we discussed the desirability of using optical data as a dependent variable. We now discuss cases where it is appropriate to use constituent data among the independent variables.

A computer doesn't know, and doesn't care, what the numbers it handles stand for; a number is a number. Numbers in a set of data that are used as calibration data for a spectrometer usually represent the optical data or some transformation thereof. However, there is no reason why the independent variables used in the calibration need be limited to the optical data.

There are two types of information that could be included along with the optical data in a calibration. The first type represents data that inherently can be represented by meaningful numerical values, usually numbers that represent real, meaningful physical quantities. Examples of such physical quantities would be temperature, pressure (accurately) known concentration of an interferent, density, etc. Information of that sort could be included with the optical information as the input to a regression program, and the results will include a regression coefficient for that variable along with coefficients for the optical data. The auxiliary statistics will then be appropriate for the calibration including that variable. If the computer program in use does not provide explicit provision for including such information, then the numbers can be put in place of the data corresponding to one item of optical data. The computer will do its job; it will then be up to the operator to keep track of the meaning of the particular variable that has been used that way.

A caveat is in order here: if a calibration is developed using extraneous information of the sort described, then the corresponding information must forever after be available for the future samples that are to be analyzed. Failure to include the extraneous information will degrade the accuracy of the calibration disastrously.

The second type of extraneous variables found particularly useful in NIR analysis, called *indicator variables*, are discussed in a separate chapter.

8.11 PRINCIPAL COMPONENTS: INTRODUCTION

Strangely, there seems to be a similarity between principal components and the weather; to paraphrase the ancient saying: "everyone talks about it, but nobody seems to know quite what it is." At least, inspection of a number of sources seems to reveal only the lack of any statement that begins "Principal components is/are ..." and then continues with a definition that is complete, concise, clear, consisting of words rather than mathematics, even though the initial mathematical foundation for what is now called *principal component analysis* (PCA) can be traced as far back as 1901 [1], and the development of the concept into its modern form is credited to Hotelling in 1933 [2].

Recent applications to NIRA are found in References 3–7, as well as a number of papers by Martens concerned mainly with partial least-squares (PLS) analysis (see also Chapter 9) where principal components are discussed peripherally, usually for comparison purposes (e.g., [8,9]).

Despite the plethora of work using this method of data analysis, the available definitions found in current literature that come closest to the ideal are as follows:

"Abstract factor analysis (also called principal component analysis) is a mathematical technique that can be applied to laboratory data to determine how many factors, or components, have contributed to an observed event" [3]. (Not very complete)

“Principal component analysis is the simplest and most commonly used bilinear method” [8]. (Not complete, and definitely not concise, considering that bilinear methods required the previous $1\frac{1}{2}$ pages to describe)

“Principal components are linear combinations of random or statistical variables which have special properties in terms of variances” [10]. (Not complete, not clear — what are those “special properties?”)

“PCA is a data reduction technique which extracts from a large number of variables, 700 $\log 1/R$ measurements for example, measured on a given set of samples, a much smaller number of new variables, ten say, which account for most of the variability between samples” [11]. (Again, not complete)

“The Principal components of a random vector X are the elements of an orthogonal transformation of X which have zero correlations” [12]. (Also incomplete)

“The basic idea of principal components analysis is to describe the dispersion of an array of n points in p -dimensional space by introducing a new set of orthogonal linear coordinates so that the sample variances of the given points with respect to these derived coordinates are in decreasing order of magnitude. Thus the first principal component is such that the projections of the given points onto it have maximum variance among all possible linear coordinates; the second principal component has maximum variance subject to being orthogonal to the first, and so on” [13]. (Almost complete, but lacks a bit in the “concise” and “clear” departments)

“Briefly, the first principal component is (any) vector (commonly taken to have length 1) such that the perpendicular projections of the points onto that vector have the largest possible variance” [14].

“Principal components analysis is a mathematically appealing tool for examining multivariate data” [9].

“Several versions of PCR exist ... but in the version considered here, p_1, \dots, p_A are defined as the normalized eigenvectors of $X'X$ corresponding to the A largest eigenvalues” [15].

8.11.1 DEFINITION AND DISCUSSION

That’s quite a list, and everybody seems to have a different piece of the definition. Part of the difficulty is that not everybody is discussing the same thing; some are discussing principal components while others are talking about PCA, which consists of methods of obtaining principal components.

The American Society for Testing and Materials recently adopted the following definition:

“Principal component analysis — a mathematical procedure for resolving sets of data into orthogonal components whose linear combinations approximate the original data to any desired degree of accuracy. As successive components are calculated, each component accounts for the maximum possible amount of residual variance in the set of data. In spectroscopy, the data are usually spectra, and the number of components is smaller than or equal to the number of variables or the number of spectra, whichever is less” [16]. (Much better)

Having investigated what’s available, let us now take a stab at putting all the pieces together and completing the sentence at the beginning of this section.

Thus, principal components are

1. The results of applying a mathematical technique
2. That generates eigenvectors
3. Corresponding to the largest eigenvalues of a covariance matrix
4. That are orthogonal (uncorrelated)
5. And unique linear combinations of which
6. Account for the largest possible amount of variance of the data
7. With the fewest such linear combinations.

Also, while not part of the definition, we will note here an important property of principal components: Since different principal components express the variations of the data that are uncorrelated, each component contains a representation of those variations of the data that *are* correlated with each other, across the wavelengths.

Now what does all *that* mean? Perhaps we can approach an answer to this question by comparing PCA to another operation that is similar but more familiar: Fourier analysis.

8.11.2 COMPARISON WITH FOURIER ANALYSIS

Mathematical theory tells us that any given arbitrary function, as long as it is well behaved (i.e., single-valued, continuous, continuous derivatives, etc.), can be approximated as closely as desired by adding together other functions; the more of these other functions we use, the better the approximation to the target function. We call the original function, the one to be approximated or regenerated, the *target function*. In spectroscopic applications, all functions of interest — the target function, and any of the various types of functions used to describe the target function — are spectra. Some may hesitate at this; it may seem to be stretching the definition of a spectrum to call the sine and cosine waves of Figure 8.8b–d spectra. However, for our purposes, it is convenient to define a spectrum (at least one relevant to optical spectroscopy) as a list of values, each value being associated with a wavelength; by this definition, each sine wave in Figure 8.8 is indeed a spectrum.

It is preferable that the functions we will use to do the approximating be orthogonal to each other; otherwise it might happen that more than one of these other functions could account for the same feature in the target function. It is also standard practice to adjust the magnitude of the approximating functions by scaling the function by a factor such that:

$$\text{Factor} = \left(\sum X_i^2 \right)^{1/2} \quad (8.21)$$

The process of making this adjustment is called *normalization* (not to be confused with the normalization used with Mahalanobis distances, as described in Chapter 15) and the factor is called the normalization factor.

For our purposes here, we will divide all such functions into two classes. One of these classes is the class of functions defined mathematically; this class includes the trigonometric functions (sines and cosines), Bessel functions, orthogonal polynomials, etc. When trigonometric functions are used, the determination of which trigonometric functions to use and how they should be weighted and combined in order to approximate or reconstruct a given target function is called *Fourier analysis*. Chapter 6 contains a full description of the use of Fourier analysis in NIR spectroscopy, but since we are going to use that technique for comparison, we present a brief discussion here also.

8.11.2.1 Fourier Analysis

As an example of how Fourier analysis works, consider Figure 8.8a–d. Part a presents a spectrum of hard red spring wheat; this is the target function that we will be reconstructing. Parts b–d present the first three lowest frequency sine and cosine waves. To emphasize the parallelism between Fourier analysis and PCA, we will introduce some nomenclature, and call these sine and cosine waves *Fourier components*, in analogy to principal components. Thus Figure 8.8b–d presents the Fourier components that we will use to approximate the target spectrum. Each Fourier component consists of two parts: a real (cosine) portion and an imaginary (sine) portion. The first Fourier component consist of one-half cycle each of the sine and cosine portions; the *n*th Fourier component consists of *n* half-cycles of each portion.

Figure 8.9a shows the real and imaginary parts of the initial portion (indeed, the first 25 points) of the Fourier transform of the target wheat spectrum. How are these obtained? There are a number

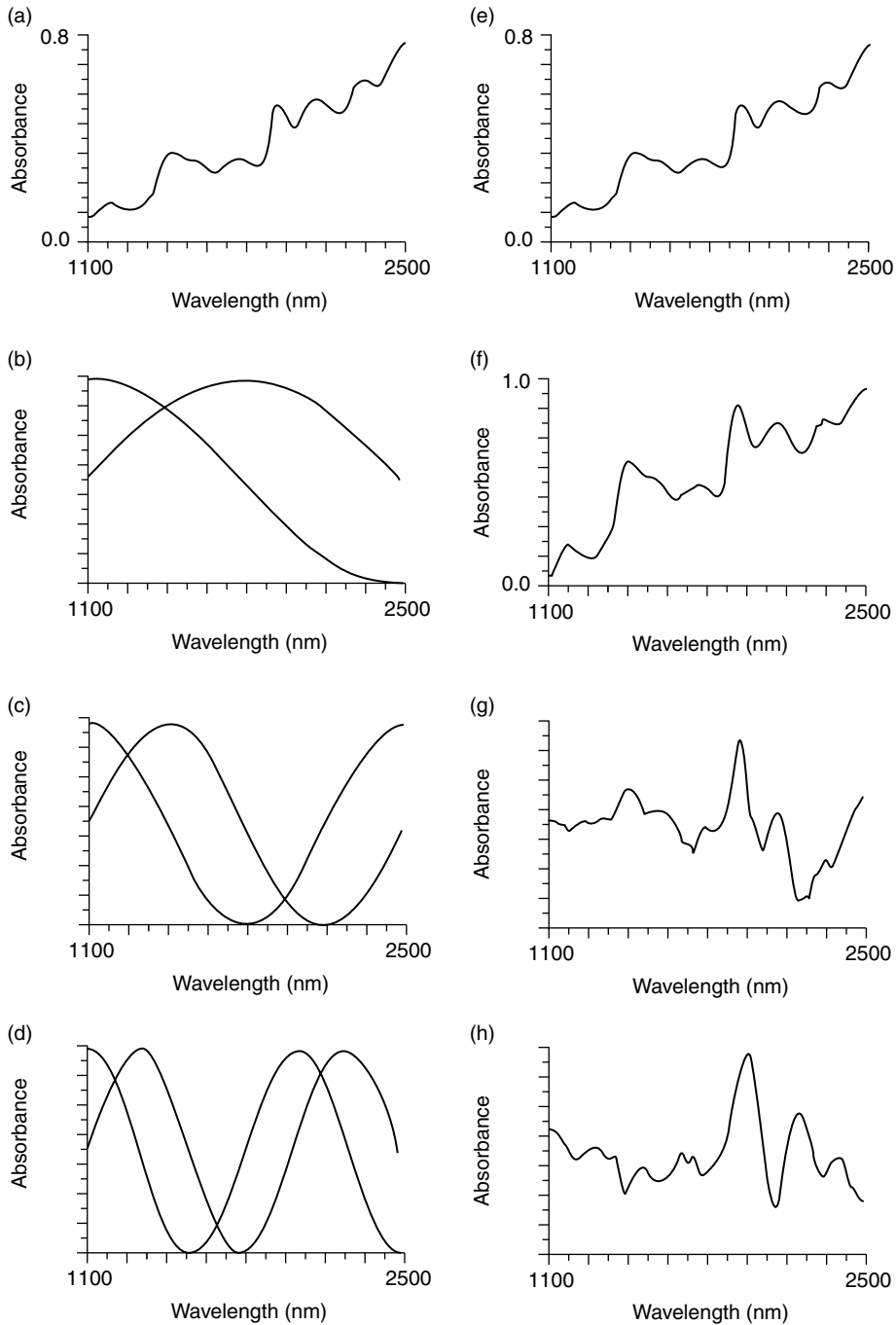


FIGURE 8.8 Target spectrum of wheat that will be reconstructed and three components of each type that will be used in the reconstruction. (a) Spectrum of wheat. (b–d) First three Fourier components. (e) Spectrum of wheat. (f–h) First three principal components of wheat.

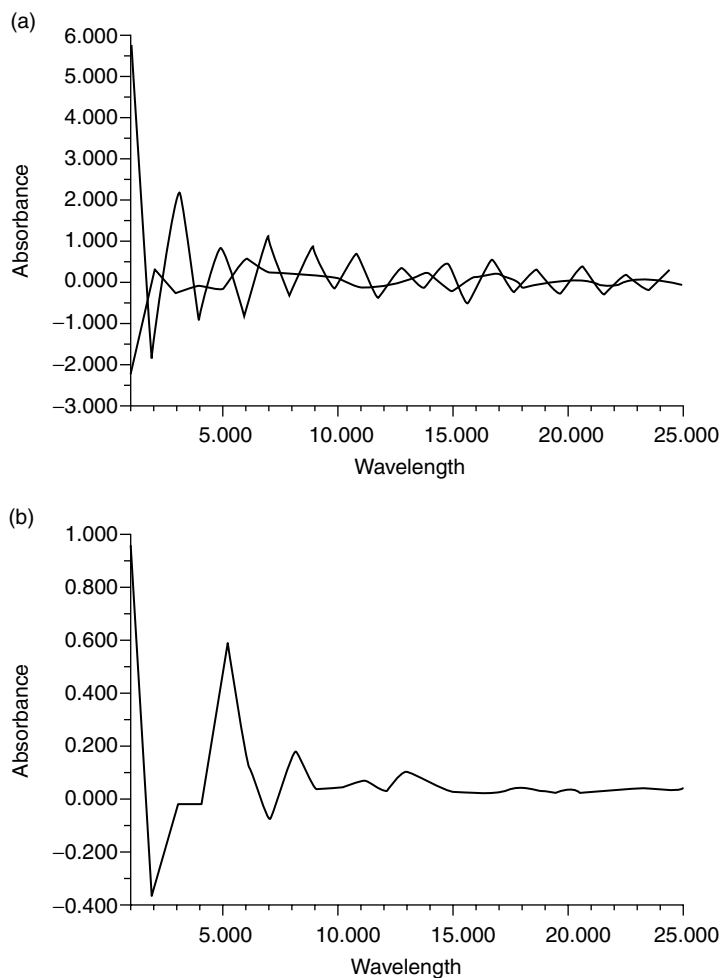


FIGURE 8.9 Transforms of the wheat spectrum. (a) Fourier transform. (b) Principal component transform.

of algorithms that can be used to obtain the Fourier transform of any given function. While the algorithm we present here is not normally the one of choice, for expository purposes the following approach is most useful. Basically each point of the Fourier transform represents the “amount” of the corresponding Fourier component in the target spectrum. Thus, if the Fourier components have been normalized as described in Equation (8.21), the first point of the Fourier transform is computed as

$$R(1) = \sum_{i=1}^m T_i C_{1i} \text{ (real)} \quad (8.22a)$$

$$I(1) = \sum_{i=1}^m T_i S_{1i} \text{ (imag.)} \quad (8.22b)$$

where $R(1)$ and $I(1)$ represent the real and imaginary parts of the first point of the Fourier transform T_i represents the value of the target spectrum at the i th wavelength, and C_{1i} and S_{1i} represent values of the cosine and sine waves comprising the real and imaginary portions of the first Fourier component

at the i th wavelength. Similarly, the n th point of the Fourier transform is computed as

$$R(n) = \sum_{i=1}^m T_i C_{ni} \text{ (real)} \quad (8.23a)$$

$$I(n) = \sum_{i=1}^m T_i S_{ni} \text{ (imag.)} \quad (8.23b)$$

Again using the nomenclature of PCA to describe a Fourier transform, the various points constituting the Fourier transform would be called the Fourier scores of the target spectrum. This computation is the same as the one by which principal component scores are obtained from principal components; hence the analogy of the names follows. The value of the i th point of the Fourier transform, in other words, the i th Fourier score, is simultaneously the result of the cross-product between the (normalized) Fourier component and the target spectrum (that is how the “amount” is determined), and the proper value by which to scale the corresponding Fourier component to regenerate the target spectrum (that is how the “amount” is used). Figure 8.10 and Figure 8.11 demonstrate this: from the Fourier transform presented in Figure 8.8a, successive approximations to the target spectrum are made by including successively more Fourier components, scaled by the corresponding values of the Fourier transform. Each approximation is overlayed on the target spectrum for comparison. Figure 8.10a–c shows the results of including first one, then two, then three Fourier components (shown in Figure 8.8b–d) in the reconstruction. Parts a–d of Figure 8.11 show the results of including 5, 10, 15, and 20 Fourier components. Note that in each case the fit shown is the *best* possible fit that can be achieved with the given number of trigonometric functions. Thus, in Figure 8.10a, the fit shown, poor as it is, is the case that can be achieved with a single sinusoid. (Note that the use of both the sine and the cosine is what creates the phase shift the produces the fit shown, which is obviously better than either one would produce alone. Hence the definition of Fourier transforms in terms of complex numbers, so as to include both real [cosine] and imaginary [sine] parts.)

For the purpose of further comparisons, Figure 8.12b–d and Figure 8.13a–d present the differences between the target spectrum and the various Fourier approximations using the corresponding number of Fourier components as Figure 8.10 and Figure 8.11. These differences represent the error of the reconstruction; as more components are included in the reconstruction process the approximation gets better and the total error decreases.

8.11.2.2 Principal Component Analysis

In the sense we have used the Fourier components, to construct an approximation to the target spectrum, principal components are just like the Fourier components. This is also demonstrated in Figure 8.8 to Figure 8.13, where corresponding operations are performed exactly in parallel with those that were done using the Fourier components.

Thus, in parts of Figure 8.8d–f we present the first three principal components of a set of spectra of hard red spring wheat; these correspond exactly to the first three Fourier components (which, as we remember, are sine and cosine functions, shown in parts a–c of that same figure) except that there is no need to represent separately real and imaginary parts of the principal components. As we are interested in them, principal components are always real. Consequently each of the principal components presented in Figure 8.8 is represented by a single curve, as opposed to the two curves needed to represent one Fourier component.

The “amount” of each principal component in any given target spectrum is determined the same way, and has the same meaning, as the amount of the various Fourier components in each target spectrum: after normalizing each principal component using Equation (8.21), the cross product between the normalized principal component and the target spectrum is computed; in principal

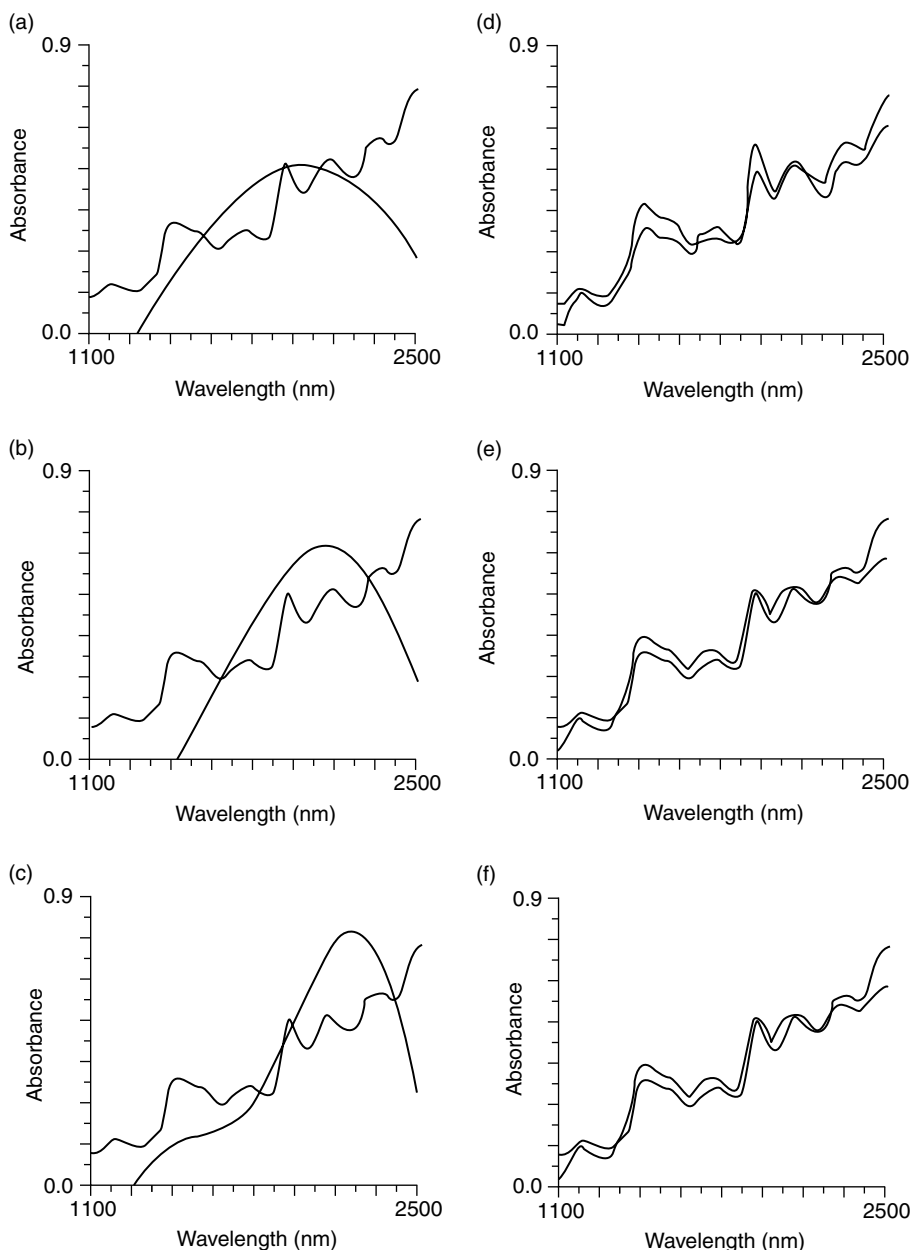


FIGURE 8.10 Reconstructions of the target wheat spectrum. (a–c) Reconstructions using one, two, and all three Fourier components, respectively. (d–f) Reconstructions using one, two, and all three principal components.

component nomenclature, this is called the principal component “score” for that principal component and the target spectrum. Computing the score for each principal component results in a series of numbers corresponding to the target spectrum, just as the cross products of the Fourier components with the spectrum resulted in the series of numbers that we called the Fourier transform of the target spectrum. Since the process of obtaining the principal component scores is identical to the process used to obtain the Fourier scores, it would seem that what we have generated by doing this a mathematical construct that we are fully justified in calling the principal component transform of

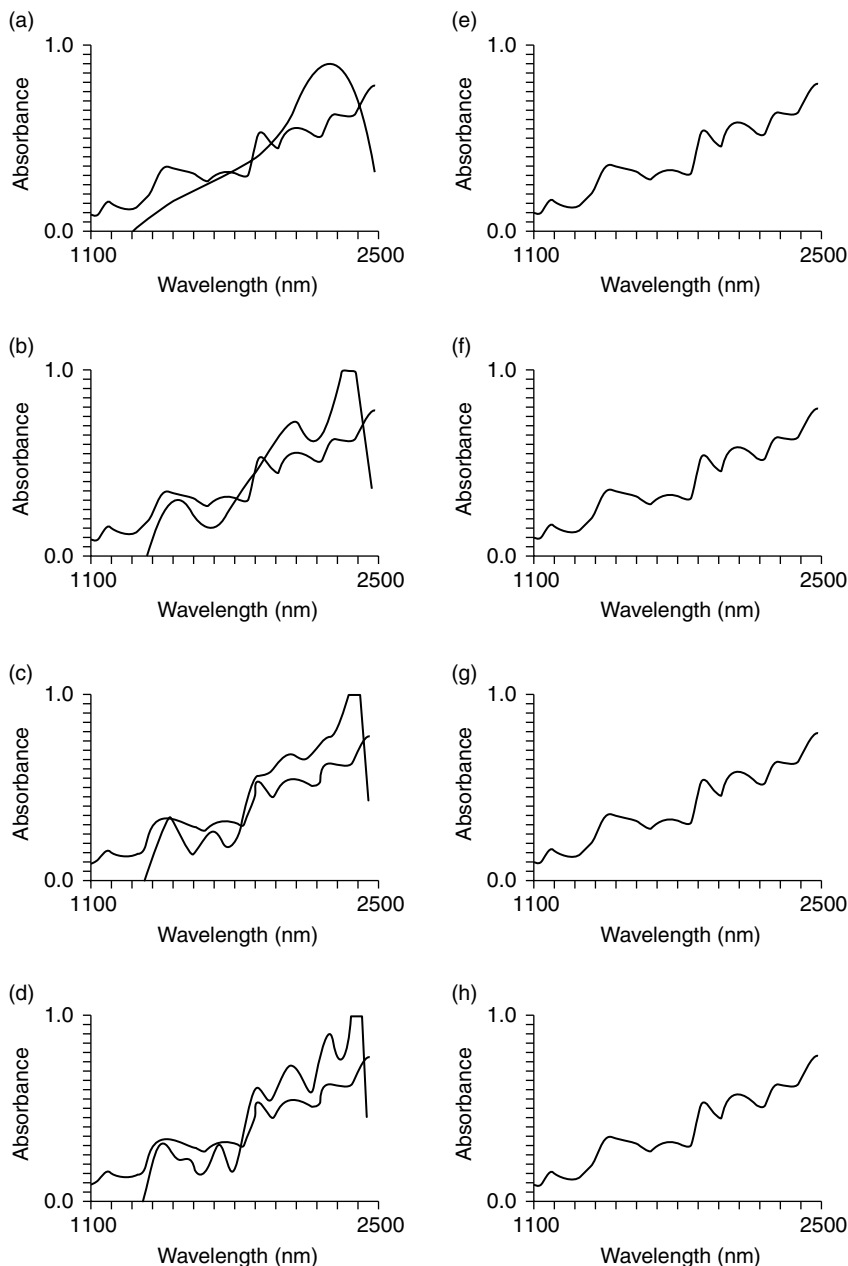


FIGURE 8.11 Reconstructions of the target wheat spectrum. (a–d) Reconstructions using 5, 10, 15, and 20 Fourier components, respectively. (e–h) Reconstructions using 5, 10, 15, and 20 principal components.

the target spectrum, just as earlier we were justified to call the elements of the Fourier transform the Fourier scores. Having recognized this parallelism between the two transform methodologies, it is clear that we can do the same things with the principal component transform as we did with the Fourier transform. For example, in Figure 8.8b, we plot the first 25 points of the principal component transform of the target wheat spectrum, exactly as we plotted the Fourier transform of that spectrum in Figure 8.8a.

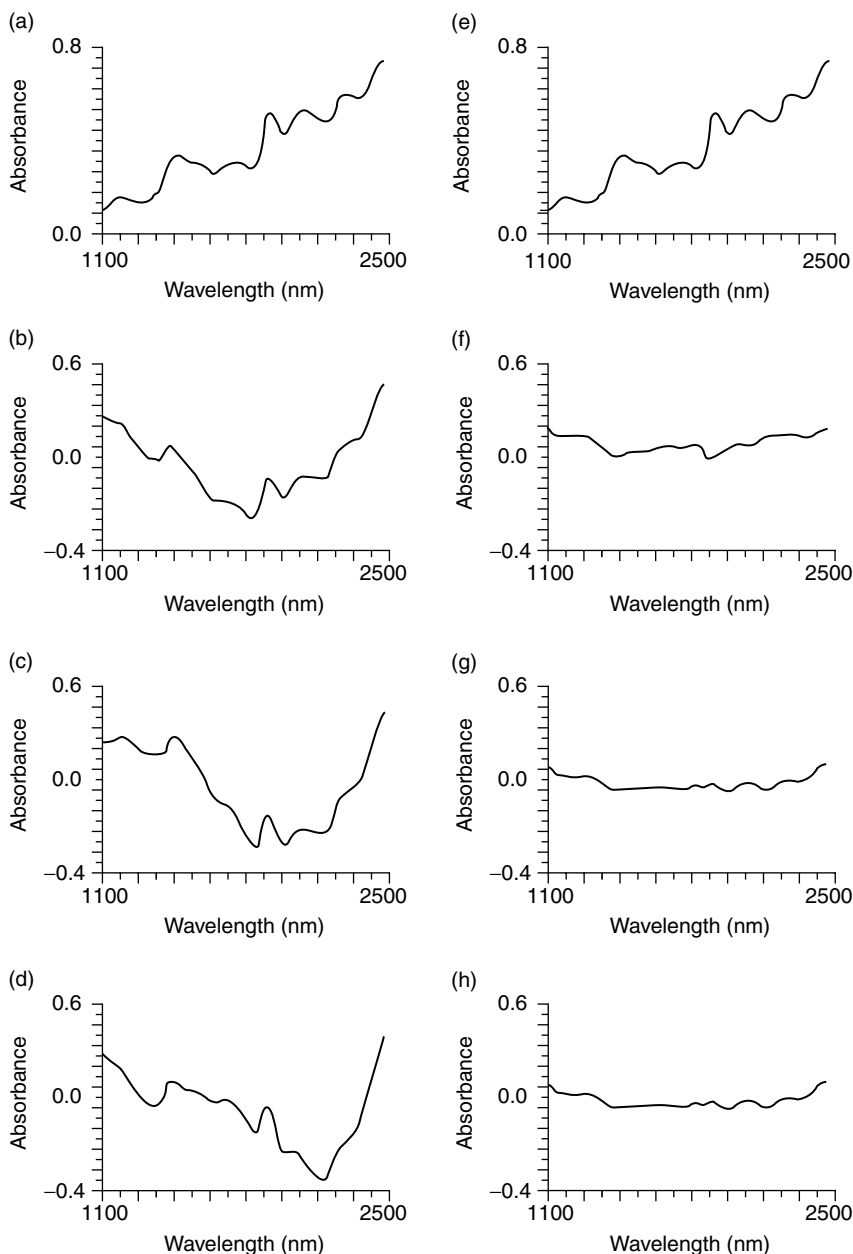


FIGURE 8.12 (a and e) Target wheat spectrum. (b–d) Differences between the target spectrum and the approximations using one, two, and three Fourier components, respectively. (f–h) Differences between the target spectrum and the approximations using one, two, and three principal components, respectively.

Similarly, in Figure 8.10d–f we plot the result of approximating the target wheat spectrum using only the first, then the first two, then the first three principal components (shown in Figure 8.8f–h) and in Figure 8.11f–h we plot the approximations using 5, 10, 15, and 20 principal components, respectively. These may be compared with the corresponding reconstructions using Fourier components shown in the first group of plots in each figure. Also, the corresponding differences from the target spectrum are presented in Figure 8.12d–f and Figure 8.13f–h, where they

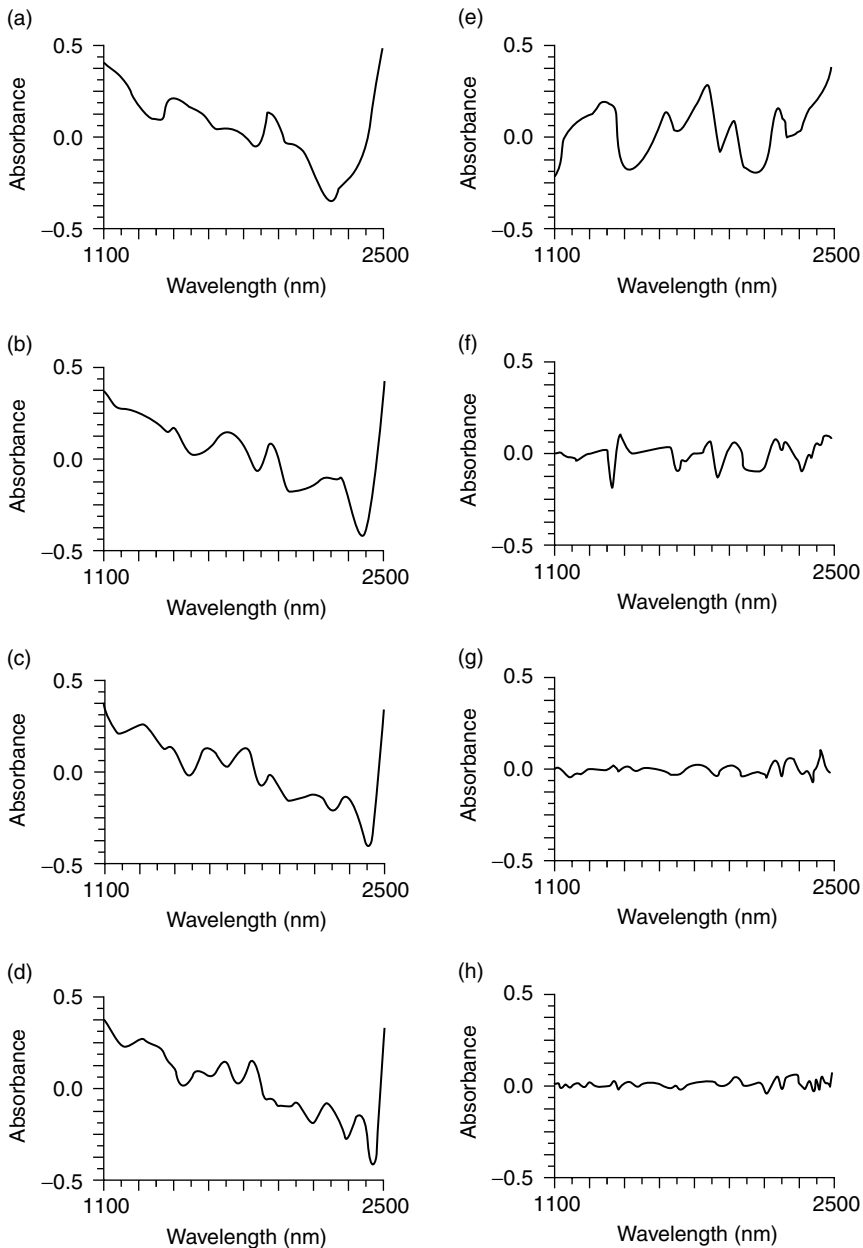


FIGURE 8.13 (a–d) Differences between the target spectrum and the approximations using 5, 10, 15, and 20 Fourier components, respectively. (e–h) Differences between the target spectrum and the approximations using 5, 10, 15, and 20 principal components, respectively. Note that the scale for the differences between the target spectrum and the reconstruction using principal components causes them to be expanded 10 times compared to the differences of the target spectrum from the Fourier components.

may be compared with the errors of the Fourier reconstruction shown in the first column of those figures.

When comparing these differences, care must be taken to note the scales of each part of the figures. All parts of Figure 8.12, except for the target spectrum shown in parts a and e, use the same scale: -0.4 to $+0.6$ absorbance. In Figure 8.13, however, while the differences of the Fourier

approximations (parts a–d) to the target spectrum is approximately the same as in Figure 8.12 (–0.5 to +0.5 absorbance), the scale for the differences from the principal component approximation (Figure 8.13e–h) expands the plot by a factor of 10 (–0.05 to +0.05).

At this point the only thing that seems to be missing that prevents us from being able to complete the picture of the parallelism between Fourier transforms (and their corresponding Fourier components) and principal component transforms (and their corresponding principal components) is information regarding the origin of the principal components that are presented in Figure 8.8.

8.11.3 WHAT MAKES PRINCIPAL COMPONENTS UNIQUE?

To begin to understand the origin of the functions that represent the principal components, we turn again to the fact mentioned in Section 8.2.2, that is, that for our current purposes we divide all functions into two classes. We also remind ourselves that the Fourier components, from which we generated the Fourier transform, are members of the class of functions that are defined by analytical mathematical expressions.

Principal components, on the other hand, are members of the other class of functions, the class of functions which contains all those functions that are defined empirically and that represent in some sense functions that are arbitrary in that they are not describable by a priori analytic mathematical expressions. Other such functions also exist. An example which is, perhaps, more familiar, is Gram–Schmidt orthogonalization. This approach requires the use of “basis functions;” these also fall into the category of being nonanalytic functions; often the basis functions are spectra of the pure materials that comprise the specimens to be analyzed, and this approach is sometimes called “curve fitting” in NIR spectroscopy. What distinguishes principal components from all the other possible arbitrary functions that can be used for similar purposes?

For the answer to this question, we go back to the definition of principal components as presented in Section 8.2.1. We noted there that principal components were orthogonal and account for the maximum possible variance of the data.

The next question is, if principal components are orthogonal, what are they orthogonal to? The answer is: to each other. Principal components are determined by computing them from the set of data that they then represent, and when principal components are computed, they are all computed in such a manner that the relation:

$$\sum X_i X_j = 0, \quad i \neq j \quad (8.24)$$

holds for any pair of principal components in the set. This condition is not unique; many sets of functions obey the relation expressed by Equation (8.24) (e.g., the various sine and cosine waves of the Fourier analysis discussed above all follow this rule), but it is one necessary condition. Essentially, it states that all principal components computed from a given dataset should be uncorrelated.

The condition that does uniquely define principal components, and distinguishes them from all other possible ways of expressing any given function as the sum of other orthogonal functions, is part 6 of the definition in Section 8.11.1: the requirement that each principal component should account for the maximum possible amount of variance in the set of data from which it is computed.

This means that if we compute a principal component from a set of spectra, as Figure 8.8f is a principal component for wheat, then we can fit that principal component to each of the spectra in the dataset from which it was computed. Figure 8.10d, for example, shows the first principal component of wheat being fit to a wheat spectrum; all that need be done is to successively consider each of the spectra in the initial set as the target spectrum. Then, scaling the principal component properly, it can be subtracted from each of the spectra in the initial set, just as we subtracted the first principal component from the target spectrum to produce the difference spectrum shown in Figure 8.12f. Having calculated all these residual spectra (“residual” referring to the difference: it is what “remains”

after the principal component is subtracted) we can then compute the following quantity:

$$\text{Residual sum of squares} = \sum_{i=1}^n \sum_{j=1}^m R_{ij}^2 \quad (8.25)$$

where the j subscript refers to all the wavelengths in spectrum and the i subscript refers to all the spectra in the dataset; that is, the residual sum of squares (RSS) is the sum of squares of all the residuals in the entire dataset.

Having performed this computation for the principal components, we could perform the same computation for the residuals obtained from any other set of functions used to fit the data, for example, the Fourier component that is displayed in Figure 8.8b, with its fit shown in Figure 8.10a and residual shown in Figure 8.12b.

Having computed and compared all these values of RSS we note the following facts: first, the total sum of squares (TSS), that is, $\sum_i \sum_j (X_{ij} - \bar{X}_j)^2$ where X_{ij} is the absorbance of the i th spectrum at the j th wavelength and \bar{X}_j is the mean absorbance at the j th wavelength) for any given set of data is constant (since it depends only on the data and not on any fitting functions); and second, the mathematical/statistical operation known as ANOVA tells us that the TSS is in fact equal to the sums of squares of each contribution. This means that the sum of squares not accounted for by the fitting function (or functions) is equal to the sum of squares of the residuals, that is, to the error of the approximation. Thus, since by their definition, principal components account for the maximum possible amount of variance (which in this case also holds for the sum of squares), the ANOVA guarantees that principal components are also distinguished by the fact that *the RSS obtained from using principal components is smaller than that obtained from any other possible set of fitting functions*, or, in other words, a principal component (specifically, the first principal component) will fit the data better than any other possible function can. Furthermore, having fit the spectra with one principal component, the residual spectra can be used to generate a second principal component which will be orthogonal to the first, *and will give the smallest RSS of all possible functions that could be used to fit the set of residual spectra resulting from fitting the first principal component*. Thus, in this sense, the principal components fit the target functions better with fewer components than any other means of trying to fit and reconstruct the target spectra, and will give a better reconstruction of the target spectra than any other functions will, using the same number of functions. Thus the first two principal components will fit the data better than any possible pair of other functions, the first three principal components will fit better than any possible triplet of other functions, etc.

As an example, we can compare the fits obtained using the Fourier components to those obtained using the principal components. Figure 8.10 to Figure 8.13 are set up so that the left-hand part of each figure is directly comparable to the right-hand part. For example, Figure 8.11b and Figure 8.11f each use ten components to reconstruct the target spectrum; the left-hand side shows the reconstruction using ten Fourier components while the right-hand side shows the result of using ten principal components.

Comparing Figure 8.10 to Figure 8.13, it is clear that in all cases the principal components give a closer fit to the target spectrum than the Fourier components do, as long as we compare the result of using one set to the result of using the *same number* of the other.

Strictly speaking, there is no guarantee that this condition will hold true for any one single spectrum; Equation (8.25) is guaranteed true only for the entire set of data. However, as Figures 8.10 to Figure 8.13 show, in practice Equation (8.25) often holds true on a spectrum-by-spectrum basis also.

Another word of caution in this regard: this characteristic of principal components, to approximate the data better than any other mathematical function that might be used for that purpose, is defined only for the dataset from which the principal components were computed. It may be expected that it will also hold true for other data of the same type as that from which the principal components were created, but that is not guaranteed, and it is certainly not so for different types of data.

8.11.4 ARE PRINCIPAL COMPONENTS BROUGHT BY THE STORK?

The foregoing is all well and good, and explains what the important characteristics of principal components are, but tells us nothing about how they are generated. Principal components are computed according to the following algorithm (or one of its variations):

Consider a set of n spectra, each containing m wavelengths. First, the m arithmetic means of the data are computed for each wavelength, then this mean spectrum is subtracted wavelength by wavelength from each of the spectra in the set. An $m \times m$ array of cross products is created from each spectrum, after the mean spectrum has been subtracted, such that the i, j th member of the array is the product of the value of the spectrum at the i th wavelength times the value of the spectrum at the j th wavelength. Corresponding terms of the arrays from each spectrum are added together; the resulting array is called the sum of cross products matrix, for obvious reasons. Mathematically, this can be expressed as

$$X\text{-prod}_{i,j} = \sum_{k=1}^n (X_{i,k} - \bar{X}_i)(X_{j,k} - \bar{X}_j) \quad (8.26)$$

The principal components are the eigenvectors of the sum of cross-products matrix; we shall neither prove this result nor discuss how eigenvectors are obtained. The former requires more advanced mathematics than is appropriate for our level of discussion; the latter is somewhat farther removed from NIRA than would warrant discussion here. Both topics are well-described elsewhere, the former in texts of mathematical statistics [10] and the latter in texts about numerical analysis [12].

However, viewed as an eigenvalue problem, principal components have some features of interest. First of all, we note that eigenvectors are solutions to the equation:

$$[V][X] = k[X] \quad (8.27)$$

where $[X]$ is the sum of cross-products matrix, $[V]$ is an eigenvector that satisfies the equality, k is a constant that along with $[V]$ satisfies the equality; and $[V][X]$ represents a matrix multiplication.

The eigenvalue (i.e., the constant in Equation (8.27)) has a useful property: The value obtained by dividing the eigenvalue associated with a given eigenvector (principal component) by the sum of all the eigenvalues is the fraction of the variance that is accounted for by that corresponding eigenvector. Since each principal component as it is computed accounts for the maximum possible variance in the data (a corollary of the fact that the RSS is the smallest possible), the first few principal components account for the maximum variance due to effects of real physical phenomena on the spectra. This is so because real phenomena will have systematic effects on the spectra; the systematic effects will, in general, be larger than the random effects due to, say, noise. Thus only noise is left for the later components; furthermore each noise component will have approximately the same magnitude. Consequently, one way to identify those principal components that represent the noise is to notice at which point the eigenvalues become nearly constant, as more principal components are computed.

8.12 CALIBRATION

It should be noted that computation of the principal components of spectroscopic data is normally based solely on the optical data; there is no explicit relationship between the principal components and the composition of the specimens comprising the set from which the spectra were measured. The usual interest of practitioners of NIR spectroscopy, on the other hand, is in methods of performing quantitative analysis using spectroscopic techniques. To do this, something more is required than just the extraction of the principal components from the data; there must be a method of relating the principal components to the constituent for which a calibration is desired.

Principal components are considered superior as predictor variables than the optical data that are directly produced by NIR instruments. This is because, since noise is rejected from the initial

principal components produced, these principal components represent true sources of variation of the spectra, presumably caused by real physical phenomena; they are sometimes called *latent variables* as opposed to the direct variables that are actually measured.

What is a latent variable? A convenient analogy here is reconstructing the spectrum of a mixture from the spectra of the pure chemical species that the mixture contains. The spectra of the pure chemicals would be the latent variables of the measured spectrum; they are latent because they are not directly accessible in the spectrum of the mixture.

Are principal components the spectra of the pure chemicals in the mixtures representing the samples? While this is possible, in general it does not happen. Principal components represent whatever independent phenomena are affecting the spectra of the specimens composing the calibration set. If one of the constituents in the specimens does in fact vary completely independently of everything else, and this constituent has a spectrum of its own, then one of the principal components will indeed represent the spectrum of that constituent. There are two problems associated with this, however. First, it is most unusual for any one constituent to vary in a manner that is exactly independent of any other. There is inevitably some correlation between the various constituents in a set of specimens; therefore any principal component will represent the sum of the effects of these correlated constituents. Second, even if independence is accomplished (say, by using a statistically designed set of specimens) there is a dependence in that the sum of all constituents must equal 100%; consequently, the principal component representing that source of independent variability will look like the *difference* between the constituent of interest and all the other constituents in the specimens. The spectrum of the constituent could be extracted from this mathematically, as we will see in the following text, but the principal component will not in itself look exactly like the spectrum of the pure constituent.

Figure 8.14 illustrates the flow of computation needed to perform a principal component calibration. Starting with a data matrix that consists of n spectra of m wavelengths each, plus the set of reference laboratory values for the constituent of interest in each specimen, the optical data are separated from the constituent values and the $m \times m$ sum of cross-products matrix computed as described in Equation (8.26).

The principal components are computed from the sum of cross-products matrix; for mathematical reasons the number of components may not be greater than the *lesser* of n or m , and may be restricted to an even smaller number, which depends on the data. Thus, the number of principal components computed may be between one and the maximum number allowed for the given dataset. We will call the number of principal components that are calculated p . The principal components then form a matrix that contains p rows and m columns; as Figure 8.8 shows, each principal component has a value corresponding to every wavelength in the original spectra.

The next step is to compute the scores for each principal component with each data spectrum as described by Equation (8.23), in other words, to compute the principal component transform for every spectrum. Thus there are p scores computed for each specimen, resulting in an $n \times p$ matrix of scores; the set of scores from each specimen is then combined with the reference laboratory value from that specimen.

It is this set of data that is used to perform the calibration calculations. Principal component regression calculations are performed exactly the same way as in a “normal” NIR calibration except the optical data are replaced by principal component scores.

The result is a set of coefficients that multiply the scores, so that predictions can be carried out using the same formalism as for normal calibration methods:

$$\hat{C} = b_0 + b_1S_1 + b_2S_2 + \cdots \quad (8.28)$$

except that, again, the variables used are the scores of the principal components included in the calibration rather than the optical data that are normally used. Except for this difference in the data used as the input to the multiple regression routine, the calibration is normal in every respect;

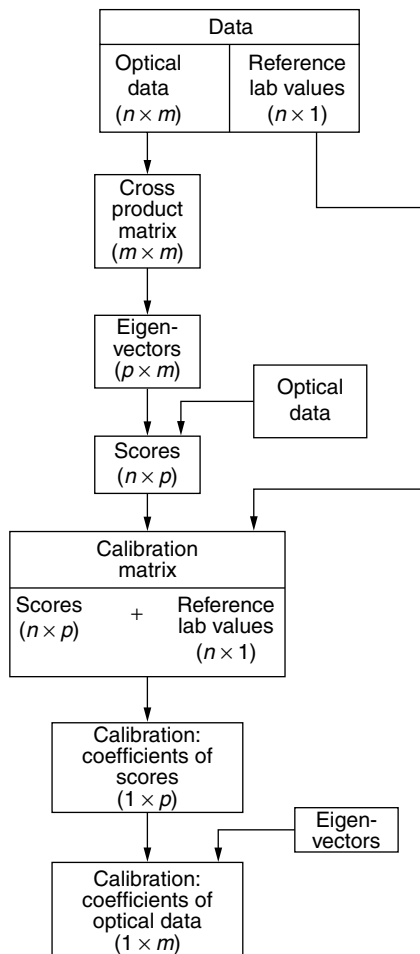


FIGURE 8.14 Flow of computation in a principal component calibration. Some files are read multiple times (e.g., the original data) and some are both created and used by the calibration program (e.g., the principal components themselves, i.e., the eigenvalues).

therefore it is possible to compute not only the calibration coefficients but also any and all of the auxiliary statistics that are associated with a calibration.

To perform prediction using this approach requires retaining not only the calibration coefficients but also the lists of values representing the principal components that were used to create the calibration, so that when unknowns are measured in the future the scores can be calculated. This is necessary because it is the scores that the calibration coefficients operate on.

However, this situation can be simplified as follows: Note that, in similar fashion to Equation (8.23), each principal component score is computed from the optical data according to the values of the corresponding principal component as follows:

$$S_i = P_{1i}X_1 + P_{2i}X_2 + \cdots \quad (8.29)$$

where X_j is the optical data for the j th wavelength and $P_{j,i}$ is the value of the i th principal component at the j th wavelength. Then, substituting the proper expression from Equation (8.29) for the various

S_i in Equation (8.28), and rearranging, we find that:

$$\hat{C} = b_0 + (b_1P_{11} + b_2P_{21} + b_3P_{31} + \cdots)X_1 + (b_1P_{12} + b_2P_{22} + b_3P_{32} + \cdots)X_2 + \cdots \quad (8.30)$$

Each item in the parentheses in Equation (8.30) is a constant; therefore each parenthesized expression is a constant and they may be replaced by individual constants, leading to an equation of the form:

$$\hat{C} = b_0 + k_1X_1 + k_2X_2 + \cdots \quad (8.31)$$

This is now an equation in which the k_i , the coefficients, are now coefficients of the optical data rather than coefficients of the principal component scores. This has three consequences: (a) It is not necessary to retain the values of the principal components. (b) Because they have the same form as normal calibration constants, these coefficients may be keyed directly into current NIR spectrometers and used for prediction in a standalone mode. (c) Because there is one of these coefficients corresponding to every wavelength in the original set of spectra, this calibration equation fits the definition of a “spectrum” given in Section 8.11.2. Therefore, it may be plotted as a spectrum or treated in any other manner that a spectrum can be treated.

For example, Figure 8.15 shows the spectra corresponding to the calibrations for protein obtained from the set of wheat data we used for our illustrative examples. These spectra correspond to the use of one, two, and three principal components to develop the calibration equation. The plot of the calibration equation can be examined to determine the characteristics of the spectra that are important to the prediction process. In Figure 8.15, for example, we see positive excursions at 2180 nm (corresponding to the absorbance of the protein amide) and a doublet at 1700 nm (corresponding to the overtone of C—H stretching); therefore these are important contributors to the protein calibration.

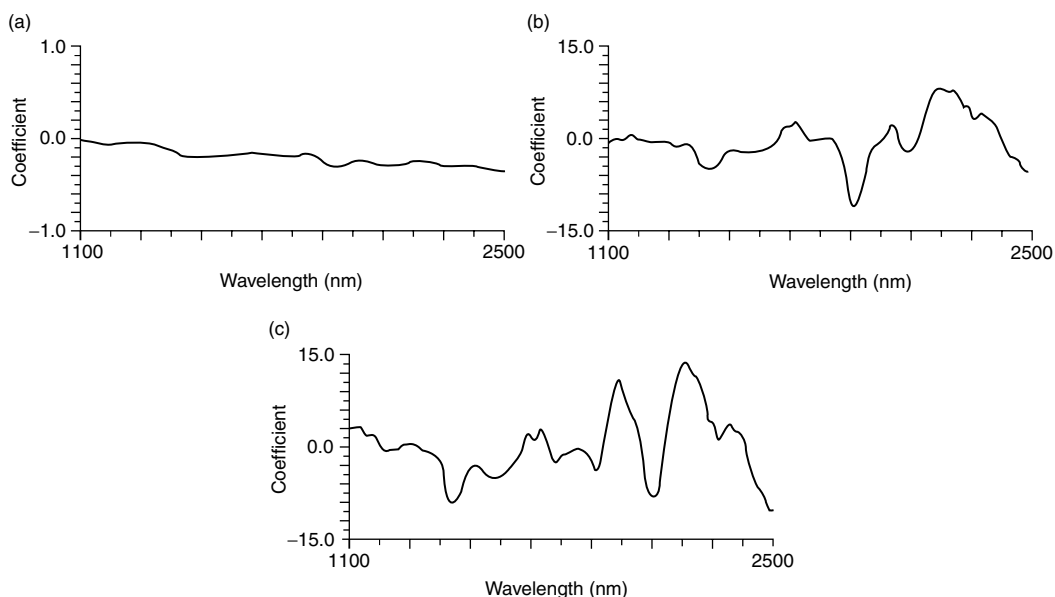


FIGURE 8.15 The calibration equation, after conversion into the form of coefficients of the optical data, has the characteristics of a spectrum, and may be plotted as such. The three calibrations shown here are all calibrations of the hard wheat specimens, of which one is shown in Figure 8.8. The calibrations differ in whether one, two, or three principal components were included in the calibration calculations (parts a–c, respectively).

On the other hand, besides some ink, we find negative excursions at 1940 nm (indicating a correction for water) and a 2100 nm (indicating a correction for starch).

8.13 CHARACTERISTICS

These are several other characteristics of principal components that make them interesting and useful for several different purposes.

In Section 8.12 we discussed the fact that principal components represent latent variables and that these are orthogonal and mathematically independent. We also noted that if a given physical phenomenon was in fact independent of the other sources of variation of the data, then a single principal component will represent the spectrum due to that phenomenon. This consideration is general in that the phenomenon need not be a chemical constituent. While it is common to think only of chemical constituents in this regard, the mathematics of the situation are general; it therefore becomes meaningful to speak of the spectrum of particle size or the spectrum of temperature in a given product, even though ordinarily we would only speak of the effect of particle size and of temperature on the spectrum of the chemical constituents.

It is not necessary that values of the phenomenon of interest be known in order to obtain its spectrum. We will see in the following text that it is possible to obtain the spectrum of a pure constituent from the spectra of mixtures when the values of concentration of the constituent in each mixture are available, but this is not necessary if the constituent concentration (or other phenomenon of interest) is uncorrelated with anything else in the training set. This being the case, this is a method of obtaining spectra of the constituents of a material even when the material is not being analyzed for those constituents.

Another aspect of the question of principal components representing spectra of independently varying constituents is that they represent the spectrum of the constituent in its natural state. An example of interest in this regard is that it is well known that the spectrum of water in natural products is not the same as the spectrum of pure water due to the effect of hydrogen bonding, dissolved organic and inorganic materials, etc. The spectrum of water obtained through a PCA will represent the water *as it exists* in the natural product. There is a major difficulty, however, in ensuring that the water is actually uncorrelated with anything else in the material. If there is a correlation, then the spectrum of the water will be contaminated with the spectrum of whatever it is correlated with; and there is a danger of assigning effects to the water spectrum that are actually due to the presence of another constituent with which the water is fortuitously correlated.

8.14 SPECTRAL RECONSTRUCTION

In the previous section we noted that there are two ways in which principal components can be used to obtain the spectra of the constituents of a set of data. In that section we discussed one of the approaches, namely, that if a given constituent is present in the calibration dataset and uncorrelated with any other phenomenon in the samples, then one of the principal components will be the spectrum of the constituent. Other algorithms to extract pure component spectra from those of mixtures in the absence of composition information have also been investigated [17].

Here we present the other side of the coin: if the constituent of interest is not completely uncorrelated with the other constituents, but its concentration is known in the various samples of the training set, then there is also an algorithm that can recover the spectrum of the constituent.

The approach used is similar to that of Honigs et al. [18], and which is further discussed in Chapter 16 of this book. The variable in Equation (8.4) of Reference 17 is analogous to the inverse of the coefficients of the Principal Component calibration, expressed in Equation (8.31). This is perhaps most clearly indicated from dimensional analysis; as pointed out in Reference 16, the variable C_{ab} has units of absorbance/concentration, while the units of the coefficients of any quantitative

calibration are concentration/absorbance. It is then straightforward to determine that the reconstructed constituent spectrum may be computed from the equation:

$$R_j = (0.01k_j)(1 - 0.01\bar{C}) + \bar{X}_j \quad (8.32)$$

where k_j represents the principal component calibration coefficient from Equation (8.31), \bar{C} is the mean constituent value, \bar{X}_j is the mean value of all the sample spectra at the j th wavelength, and a factor of 0.01 is required to correct the equation for the fact that constituent concentrations are normally expressed in percentages.

The use of this algorithm for spectral reconstruction has most of the same characteristics as the previous algorithm [18], including the fact that it works best on data that produce good quantitative calibration results, that it produces a spectrum representing what the “pure” constituent spectrum would be like, and that the reconstructed spectrum represents the constituent as it exists in the product. However, here we must also heed the warning, as pointed out in Section 8.13, concerning the distortion of the spectrum by fortuitous correlations. These are operative using this method of spectral reconstruction, as with any other.

The current approach, however, has the advantage that rather than being based on an individual wavelength-by-wavelength reconstruction, it used the correlated structures inherent in the dataset, and that are found by the PCA. Furthermore, if a given principal component is found not to assist the calibration process, then it can be deleted from the spectral reconstruction process also; these two facts cause it to provide more accurate reconstructions.

This is indicated in Figure 8.16, where the reconstructed spectrum of wheat protein is presented. Three reconstructions are shown using several combinations of principal components. It is clear that the reconstruction is improved with the inclusion of more components.

Another aspect to this method of spectral reconstruction is that it is not necessary to use principal components to perform the reconstruction. Any set of functions that can be used to perform a calibration can also be used in the reconstruction process. In line with the discussion in the early part of this chapter, Figure 8.17 illustrates the use of Fourier components to perform spectral reconstruction.

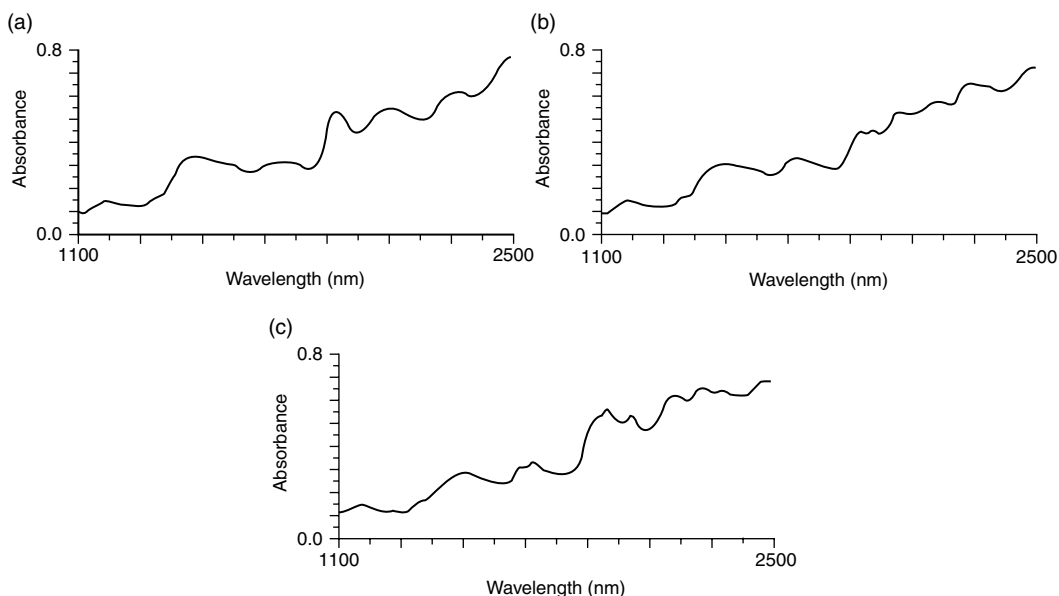


FIGURE 8.16 Spectral construction of wheat protein. (a) Reconstruction using one principal component. (b) Reconstruction using two principal components. (c) Reconstruction using three principal components.

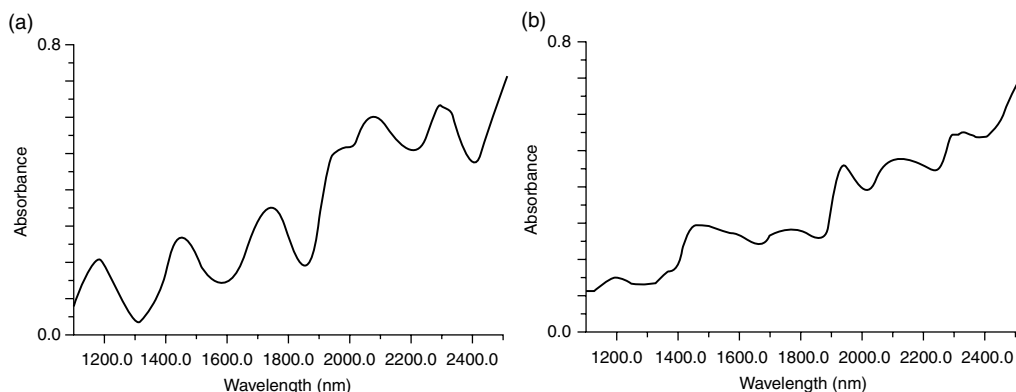


FIGURE 8.17 Spectral reconstruction of wheat protein using Fourier components. (a) Reconstruction using ten Fourier components. (b) Reconstruction using 25 Fourier components.

In this case also, use of more components result in a more accurate reconstruction. However, just as principal components account for most of the variance of the dataset with fewest components, so also does the reconstruction converge faster with principal components than with any other type of function. Comparing Figure 8.16 and Figure 8.17, it is clear that it requires 25 Fourier components to do only slightly better than one principal component in recreating the spectrum of wheat protein.

8.15 ADVANTAGES

8.15.1 LACK OF REQUIREMENT FOR WAVELENGTH SELECTION

One of the main advantages of using the principal component approach to calibration is that there is no need to perform a wavelength search. One of the great difficulties in generating NIR calibrations has been the question of selecting the wavelengths at which to calibrate; in Section 8.9 of this chapter we saw that the noise of the spectra causes large random changes in the wavelengths actually picked, whereas in actuality many wavelengths in the neighborhood of spectral peak are equivalent. This problem is exacerbated by the fact that some of the important effects superimposed on NIR spectra have no distinct absorbance bands (there is no absorbance band for particle size, for example), yet it is clear that *some* wavelength must be used in order to make a correction for this phenomenon.

The use of principal component calibrations sidesteps that entire question; as Figure 8.15 shows, the same number of wavelengths are used in the calibration regardless of the number of principal components that were included in the calculations that gave rise to the calibration.

One desirable consequence of this is that while all the wavelengths are included in the calibration equation, the difficulties usually encountered with this situation, such as the calibration coefficients “blowing up” due to overfitting, do not occur. At least, they do not occur solely on account of the use of all available wavelengths. It is still possible to overfit a calibration by including an excess number of principal components in the calibration, with results equivalent to what is obtained when too many wavelengths are used. In general, however, it is easier to avoid including too many principal components because, due to the orthogonality of the components, the values of the *t* statistics for the coefficients of the principal components are directly interpretable as indicating which components should be retained and which ones deleted from the calibration.

Furthermore, the use of principal components to create calibration equations automatically gives the proper weighting to each wavelength, which takes into account the different signal-to-noise ratios at the different wavelengths. In NIR data the signal-to-noise ratio is particularly difficult to assess. While the noise levels are very low, they are nonzero and relatively easy to measure. The signal, on the other hand, is extremely difficult to determine because the “signal” that is operative is the

change in absorbance at each wavelength due to the changing concentration of each constituent in the material. Thus, not only does each constituent have its own signal-to-noise ratio, it is virtually impossible to tell what that is for any given constituent at any specified wavelength.

The avoidance of a need to select wavelengths, however, is replaced by a corresponding requirement: the need to select the principal components to include in the calibration. However, it was previously shown that this problem is much easier to solve. The orthogonality of the principal components makes the use of the t statistic calculated for each coefficient of the principal component scores an eminently efficacious method of selection [7].

8.15.2 ROBUSTNESS

The orthogonality of the principal components combined with the fact that collectively they account for the maximum amount of variance in the data makes the use of principal components an inherently robust method of calibration. It is well known that since the principal components are orthogonal, calibrations based on them obviate the problems associated with calibrations based on the highly correlated optical data.

8.15.3 INTERPRETATION OF CALIBRATION EQUATIONS

In Section 8.12 we noted that the calibration equation is itself a spectrum, and as such it can be plotted and spectroscopic interpretations made based on the spectrum of the calibration.

8.16 QUALITATIVE ANALYSIS (SIMCA)

SIMCA is a multivariate technique of analyzing data that has only recently been implemented NIR analysis. It is fundamentally a combination of methods that individually have been used for NIR analysis for a considerable time. The acronym SIMCA stands for *soft independent modeling of class analogies*.

The basic approach is well described by Sharaf et al. [19]. While these authors point out that the SIMCA approach can be applied to a more general class of problems than simple classification (i.e., identification), approaching the method from the point of view of the more limited application will perhaps clarify the approach.

A PCA is performed on a set of data for which qualitative analysis is desired. After computing the principal components, the scores are calculated, just as described in Section 8.11.2.2.

These principal component scores are now used to perform qualitative analysis by surrounding each region of multidimensional space containing the scores corresponding to each group with a surface. While Sharaf et al. illustrate the process using parallelograms and their higher dimensional analogs, conceptually the enclosing surface could as well be an ellipse or ellipsoid, just as was done to describe the space containing the original data in Chapter 15. Thus, computing Mahalanobis distances based on principal component scores incorporates the advantages of the principal component approach to data analysis into the qualitative analysis field also.

Thus, the combination of PCA and discriminant analysis, both of which are techniques that have already found use in NIRA individually, could be combined to form still another method of greater power and wider applicability.

8.17 DATA COMPRESSION

Since the computation of principal components packs the most information into the fewest numbers, this would seem to be the ideal method of data compression, far more compact, for example, than the use of Fourier transforms provides. This is both true and false. While it may be possible to recreate

a spectrum by specifying many fewer principal component scores than Fourier scores, for example, in the case of the Fourier approach to data compression the Fourier components that will be used to do the actual recreation of the spectrum are known mathematical functions.

In the case of principal components, on the other hand, the fact that we need to specify fewer scores can be greatly overshadowed by the fact that we also need to carry around the excess baggage of the definition of the principal components themselves. If only one or a few spectra are to be recreated, then the definitions of the principal components may completely overwhelm the smaller number of scores needed. If many spectra are involved, the savings in storage space due to the smaller number of scores for each spectrum more than compensates for the overhead of having to retain the definitions of the principal components.

8.18 WHAT PRINCIPAL COMPONENTS WON'T DO

Last but not least, we want to make sure that we do not leave misconceptions in the minds of our readers. It was noted in Section 8.3 of that one of the fundamental assumptions of regression analysis is that there is no error in the independent variables; all the error should be in the dependent variables. In NIRA, this is usually the case; the reference laboratory error against which the calibrations are performed are normally the limiting error source of the calibration process.

This being the case, it should be clear that no manipulations of the independent variables, in other words, the optical data, no matter how sophisticated, can improve the calibration beyond that limit. Therefore, while the use of principal components does bring advantages with it to the calibration process, those advantages will in general *not* include obtaining a better calibration.

Now I will back-pedal a bit. In a particular case, it may happen that principal components will produce a better calibration than the standard multiple regression approach: there appears to be a contradiction, but that is indeed only appearances.

The case here is that the regression may not have been properly performed. If the principal components can produce a certain result, then the standard regression approach is also capable of producing an equivalent result. The difference is not the ultimate capabilities of the data or of the algorithm. Rather, it is that the degree of difficulty of obtaining the optimum result and the required skill on the part of the data analyst are both much higher for the standard regression case; and this is one of the true benefits of using principal components.

In Section 8.15.2, we repeated the generally held belief that since principal components are orthogonal and tend to reject noise in the data, they are more robust than other methods of analyzing data. It is certainly true that when a principal component calibration is performed, changing the number of principal component scores included in the calibration does not change the regression coefficients of those scores that are included, while in contrast, it is well known that due to the intercorrelations among the readings at different wavelengths, calibrations based directly on NIR optical data are very variable, the coefficients of the data changing considerably as different wavelengths are included in the calibration.

However, it is a truism that all statistics that are calculated from data containing a random component themselves have a random component. Thus, the mean of a given set of data that contains random noise is some particular value; that value includes the random contribution of the noise. If another set of data is collected under conditions as nearly identical to the first as possible, the mean will not be identical to the first because the contribution due to the random component will not be the same.

There seems to be no reason to believe that principal components are immune to this phenomenon. Certainly, when real samples are run via NIRA, there are enough random contributions to the data to go around. Instrumental (electronic) noise and sample noise (e.g., particle size, repack effects, nonhomogeneity) should introduce considerable variation in the data; it would be surprising indeed if these had no effect whatsoever on the principal components that are computed.

Multivariate methods of data analysis are touted as the mathematical equivalent of “wonder drugs;” they are supposed to cure all ills. Indeed, they go a long way in that direction. However, the side effects of their use have not been widely investigated; what preliminary results exist [20] indicate that the whole story is not yet known; there is considerable variation in the principal components that are computed when different sets of nominally the same noise are added to the same initial set of data.

A corollary to the fact that the reference laboratory error dominates NIR calibration error (a condition that is right and proper, and that conforms to the theoretical basis of regression analysis) is the fact that direct comparisons of calibration results using different mathematical transformations of the data are generally invalid. The reason for this is the fact that the statistic available for the comparison, the sum of squares due to deviation from the regression (or the corresponding mean square, which is normally reported as the SEE), is essentially the same in the different cases because it reflects only the reference laboratory error. Variations due to changes in any of the other error sources are swamped out by this one source. Thus, it is not ordinarily possible to tell whether a given change in the magnitude of error due to one of these other sources is real or simply an artifact of using a different set of data, with different noise characteristics, simply by comparing values of the SEE or of the mean square.

The situation is not hopeless; however, it does require that a more complicated approach be used. In particular, it is necessary that suitable data be collected that will allow a direct calculation of the error due to any particular source. This in turn requires that data should be collected according to a statistically designed scheme, so that the source of variation to be studied can be separated from the others. One example of how this approach can be applied to NIR data to extract the values of several sources of variation at the same time has appeared in the literature [7]. Statistical texts are available that describe a number of other experimental designs that may be applied. For unusual circumstances it may be necessary to consult a statistician; such a consultation is a good idea anyway, as statistical methodologies are fraught with pitfalls to trap the unwary and inexperienced, and an expert guide is invaluable.

There is an additional difficulty in determining which principal components represent systematic (i.e., real) phenomena and which represent the noise of the measurements. This is important because the number of principal components that represent real phenomena is the effective rank of the matrix. Previously, the eigenvalues resulting from the PCA have been studied with a view to trying to determine which are significant.

Another approach has also been presented [6]. Noise, being random, will change when subjected to an averaging process. Since principal components representing real phenomena are purely systematic, they will not change when averaged. Thus, using a statistically designed set of data, where principal components can be calculated on both averaged and unaveraged data, will allow a determination of which ones are systematic and which are random, since only the random ones will change between the two cases. The coefficients of the principal component scores upon performance of the calibration calculations will allow for an easy way to compare the two cases, component by component.

ACKNOWLEDGMENT

The author gratefully acknowledges permission of Bran and Luebbe Analyzing Technologies to include copyrighted material originally appearing in the PCA section of the IDAS-PC manual.

REFERENCES

1. K. Pearson, *Philos. Mag.* (Ser. 6), 2: 559–572 (1901).
2. H. Hotelling, *J. Educ. Psychol.*, 24: 417–441, 498–520 (1933).
3. I. A. Cowe and J. W. McNicol, *Appl. Spectrosc.*, 39: 257–266 (1985).

4. I. A. Cowe, J. W. McNicol, and D. C. Cuthbertson, *Analyst*, 110: 1227–1232 (1985).
5. I. A. Cowe, J. W. McNicol, and D. C. Cuthbertson, *Analyst*, 110: 1233–1240 (1985).
6. H. Mark, ChimicaOGGI, edizione: Teknoscienze Srl. (Sept.): 57–65 (1987).
7. H. Mark, *Anal. Chem.*, 58: 2814–2819 (1986).
8. H. Martens, S. Wold, and M. Martens, A Layman's Guide to Multivariate Analysis, in *Food Research and Data Analysis* (Harald Martens and Helmut Russwurm, eds.), Applied Science Publishers, New York, 1983, p. 482.
9. G. P. McCabe, *Technometrics*, 26: 137 (1984).
10. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, 1958, p. 282.
11. B. G. Osborne and T. Fearn, *Near Infrared Spectroscopy in Food Analysis*, John Wiley & Sons, New York, 1986, p. 108.
12. W. J. Kennedy and J. E. Gentle, *Statistical Computing*, Marcel Dekker, New York, 1980, p. 566.
13. R. Gnandesikan, *Methods for Statistical Data Analysis of Multivariate Observations*, John Wiley & Sons, New York, 1977, p. 8.
14. K. Enslein, A. Ralston, and H. S. Wilf, *Statistical Methods for Digital Computers*, John Wiley & Sons, New York, 1977, p. 306.
15. G. L. Durst, *Am. Lab.*, August: 116 (1986).
16. Standard Definitions of Terms and Symbols Relating to Molecular Spectroscopy, ASTM Vol. 14.01, Standard E131-90.
17. J. Liu and J. L. Koenig, *Anal. Chem.*, 59: 2609–2615 (1987).
18. D. Honigs, D. M. Hieftje, and T. Hirschfeld, *Appl. Spectrosc.*, 38(3): 317–322 (1984).
19. M. A. Sharaf, D. Illman, and B. R. Kowalski, *Chemometrics* (Volume 82 in Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications), John Wiley & Sons, New York, 1986, p. 242.
20. H. Mark and J. Workman, *The Effect of Principal Component Analysis in Regard to the Noise Sensitivity of Calibration Equations*, Paper #030, Pittsburgh Conference and Exposition, Atlantic City, NJ, March 9–13, 1987.

9 Data Analysis: Calibration of NIR Instruments by PLS Regression

Hans-René Bjørsvik and Harald Martens

CONTENTS

9.1	Introduction	189
9.2	Notation	190
9.3	Experimental	190
9.3.1	Equipment and Software Requirements	190
9.3.2	Laboratory Procedure.....	191
9.4	Data Analysis Method	191
9.4.1	The PLS Regression Method.....	191
9.4.1.1	Bilinear Modeling.....	193
9.4.1.2	PLS1 Regression	195
9.4.1.3	Calibration and Prediction	195
9.4.1.4	Validation and Assessment	196
9.4.2	The PLS1 Algorithm.....	198
9.5	Results and Discussion	198
9.5.1	Residual Statistics	198
9.5.2	PLSR Model	199
9.5.3	Prediction of Analyte in New Objects	203
9.6	Discussion.....	203
	Acknowledgments	205
	References	205

9.1 INTRODUCTION

Partial least-squares regression (PLSR) is a multivariate data analytical technique designed to handle intercorrelated regressors. It is based on Herman Wold's general PLS principle [1], in which complicated, multivariate systems analysis problems are solved by a sequence of simple least-squares regressions.

The PLSR method to be presented here is formally defined as a predictive two-block regression method based on estimated latent variables. The present orthogonalized multifactor version [2] was primarily developed by Svante Wold. The so-called PLS1 version of this PLSR (one single analyte) will be described in detail in this chapter. It was first applied successfully to near-infrared (NIR) data by Martens and Jensen [3], and their algorithmic representation will be used here, although with a more recent notation.

The purpose of the multivariate calibration here is to predict an analyte's concentrations y_i in objects $i = 1, 2, \dots, I$ from a set of optical densities x_{ik} at wavelength channels $k = 1, 2, \dots, K$ via a linear predictor equation shown as follows:

$$\hat{y}_i = b_0 + \sum_{k=1}^K x_{ik} b_k \quad (9.1)$$

Based on data $\mathbf{y} = (y_i, i = 1, 2, \dots, I)^T$ and $\mathbf{X} = (x_{ik}, i = 1, 2, \dots, I; k = 1, 2, \dots, K)$ from a set of I calibration samples (calibration objects), the parameters in the predictor equation are to be estimated statistically. Once obtained, this predictor equation can be used for converting X data into y estimates in new future objects.

The PLSR is a bilinear regression method that extracts a small number of "factors," \mathbf{t}_a , $a = 1, 2, \dots, A$ that are linear combinations of the KX variables, and use these factors as regressors for \mathbf{y} . In addition, the X variables themselves are also modeled by these regression factors. Therefore outliers with abnormal spectra \mathbf{x}_i in the calibration set or in future objects can be detected.

What is special for the PLSR compared to, for example, the more well-known statistical technique of principal component regression (PCR) is that the y variable is used actively in determining how the regression factors \mathbf{t}_a , $a = 1, 2, \dots, A$ are computed from the spectra X . Each PLSR factor \mathbf{t}_a is defined so that it describes as much as possible of the covariance between \mathbf{X} and \mathbf{y} remaining after the previous $a - 1$ factors have been estimated and subtracted.

To illustrate the PLSR method, we shall examine data from a laboratory experiment performed by participants in a course in quantitative chemometrics. The course was held by the authors at Østfold College of Engineering, Department of Chemistry and Chemical Engineering, Norway, Autumn 1987. The purpose of the experiment was to determine ethanol in mixtures with two similar compounds (methanol and *n*-propanol) by fiberoptic NIR spectrophotometry. Mixtures of the three different alcohols were prepared, where the concentrations were varied between 0 and 100% of each of the three alcohols. Two different mixture sets were made. One of them will be used as a calibration set and the other as a test set, in other words, mixtures where the concentration of the analyte are "unknown."

In the present illustration ethanol is regarded as the analyte to be calibrated for. The two compounds methanol and *n*-propanol will be treated as "unidentified interferents." Hence, only the ethanol concentration data (y) will be used in the data analysis.

To illustrate how the model building and interpreting of an PLSR calibration model proceeds, the experiment will be worked through step by step, and various plots will be used to explain how the data can be analyzed and interpreted.

9.2 NOTATION

In this chapter the following notation convention is used: matrices and vectors are in boldface letters, for example, \mathbf{X} and \mathbf{y} ; thus uppercase letters are used when there is more than one variable, and lowercase letters when there is only one variable. Scalars are denoted by ordinary letters, for example, q_a .

All vectors are column vectors unless written as a transpose, for example, \mathbf{x}_i^T . The symbol, for example, \bar{y} , means the statistically estimated value of the corresponding variable.

In Table 9.1 all variables used in this chapter are described.

9.3 EXPERIMENTAL

9.3.1 EQUIPMENT AND SOFTWARE REQUIREMENTS

The NIR absorbance spectra $\log(1/T)$ of the three alcohols were measured with a Guided Wave Model 200-40 spectrophotometer, equipped with germanium detector, 600 l/mm grating, Vis-NIR 6-to-1

TABLE 9.1
Notation Used

Symbol	Explanation
\mathbf{X}	Matrix containing NIR spectra $\mathbf{X} = (x_{ik}, i = 1, 2, \dots, I; k = 1, 2, \dots, K)$
\mathbf{y}	Vector containing ethanol concentration ($y_i, i = 1, 2, \dots, I$)
i	Index variable for number of objects ($i = 1, 2, \dots, I$)
k	Index variable for number of X variables ($k = 1, 2, \dots, K$)
a	Index variable for number of factors ($a = 1, \dots, A$)
\mathbf{t}_a	Score vector for factor a
\mathbf{p}_a^T	Loading for factor a
\mathbf{w}_a^T	Loading weight for factor a
$\bar{\mathbf{x}}$	Mean of X over all objects and for each variable $k = 1, 2, \dots, K$
\bar{y}	Mean of y over all objects
b_0	"Intercept"
\mathbf{b}_k	Vector containing regressors
\mathbf{E}_a	Residual in X matrix after a factors
\mathbf{f}_a	Residual in y vector after a factors
RMSEP	Root mean squares of error prediction
PLSR	Partial least-squares regression
PCR	Principal component regression
SMLR	Stepwise multiple linear regression
PLS1	PLSR with only one y -variable

probe, and a 5 mm probe tip. The total optical fiber length was about 4 m and the optical pathlength of the sample was 10 mm. The data acquisition (instrument controller) and analysis were done on a Copam AT (IBMAT compatible computer). The calibration analysis was done in an early version (2.0) of the Unscrambler package, (www.camo.com) originally developed by the second author (described in Unscrambler user manual) [4]. The final graphics plots were made in the Grapher program (Golden Software, Inc., Golden, Colorado), described in the Grapher reference manual [5].

More recent information about multivariate calibration may be found in Reference 6. An updated introduction on how multivariate NIT calibration can be seen as part of a general chemometrics/qualimetrics method is given in Reference 7.

9.3.2 LABORATORY PROCEDURE

A calibration set of 15 ethanol/methanol/*n*-propanol mixtures and a similar test set of 11 such mixtures were prepared by mixing pure compounds. Their weight fractions are given in Table 9.2 and Table 9.3, respectively.

The alcohols that were used in this experiment were of analytical qualities and are produced by E. Merck (methanol), A/S Vinmonopolet (ethanol), and E. Merck (*n*-propanol).

For each of the mixtures (in total 26 objects) the NIR absorbance spectra was recorded between 1100 and 1600 nm at every 5 nm, in total 101 X variables.

Figure 9.1 shows the NIR absorbance spectra of the three pure alcohols, the analyte ethanol E in solid line and the two "unidentified interferents" methanol (M) and *n*-propanol (P) in dashed lines.

9.4 DATA ANALYSIS METHOD

9.4.1 THE PLS REGRESSION METHOD

Partial least squares is a loose term for a family of philosophically and technically related multivariate modeling methods derived from Herman Wold's basic concepts of iterative fitting of bilinear models

TABLE 9.2
Calibration Set of the Three Alcohols: Methanol,
Ethanol, and *n*-Propanol

Sample no.	Analyte % ethanol	Interferent % methanol	Interferent % <i>n</i> -propanol
1	0.0	0.0	100.0
2	0.0	100.0	0.0
3	100.0	0.0	0.0
4	33.0	33.4	33.6
5	24.9	49.8	25.3
6	25.3	24.9	49.8
7	50.0	25.0	25.0
8	0.0	50.1	49.9
9	33.4	66.6	0.0
10	33.4	0.0	66.6
11	0.0	75.0	25.0
12	0.0	25.0	75.0
13	66.6	33.4	0.0
14	66.7	0.0	33.3
15	33.0	33.4	33.6

TABLE 9.3
Prediction Set of the Three Alcohols: Methanol,
Ethanol, and *n*-Propanol

Sample no.	Analyte % ethanol	Interferent % methanol	Interferent % <i>n</i> -propanol
1	0.0	100.0	0.0
2	100.0	0.0	0.0
3	0.0	0.0	100.0
4	50.0	50.0	0.0
5	0.0	50.0	50.0
6	49.9	0.0	50.1
7	0.0	75.0	25.0
8	0.0	25.7	74.3
9	25.0	50.0	25.0
10	50.0	25.0	25.0
11	25.0	25.0	50.0

in several blocks of variables. These concepts arose around 1975 as a practical solution to concrete data analytical problems in econometrics: When multivariate statistical modeling tools based on maximum likelihood principle are applied to real data with many colinear variables and a modest number of observations, grave identification and convergence problems were often encountered. These problems were overcome by the more empirical PLSR approach, based instead on the least-squares principle.

The PLS principle can be applied to a variety of modeling problems. The most common implementation in econometrics has been one-factor path modeling of multiblock relationships. These relationships can be estimated as either formative models or reflective models. We here

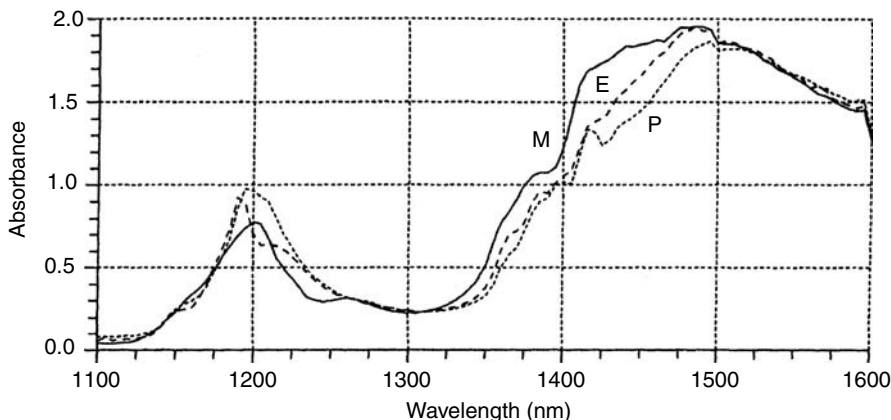


FIGURE 9.1 NIR absorbance spectra of ethanol (E) in solid line, methanol (M), and *n*-propanol (P) in dashed lines.

concentrate on reflective two-block linear (X , Y) formulations. Measured variables X , $Y = f(\text{latent variables}) + \text{residuals in } X, Y$.

9.4.1.1 Bilinear Modeling

PLSR belongs to the class of bilinear regression methods, which again belongs to a wider class of regressions on estimated latent variables.

The purpose of using PLSR in multivariate calibration is to obtain good insight and good predictive ability at the same time. The responses of an NIR instrument at different wavelengths $\mathbf{X} = (x_1, x_2, x_3, \dots, x_K)$ are usually highly collinear (intercorrelated), especially so for scanning NIR instruments. A problem therefore arises in ordinary multiple linear regression (MLR), since MLR assumes each X variable to have unique information about y . To be useful, an NIR calibration method has to handle the phenomenon of multicollinearity.

In classical stepwise multiple linear regression (SMLR) the collinearity is handled by picking out a small subset of individual, distinctly different X variables from all the available X variables. This reduced subset is used as regressors for y , leaving the other \mathbf{X} variables unused. In analogy, derivative and ratioing calibration techniques handle the collinearity by selecting a small subset of X variables in certain fixed-difference and ratio patterns, to be used as regressors.

In regression methods employing estimated latent variables we seek to treat the collinearity in the X data as a stabilizing advantage rather than as a problem. This is done by seeking to compress all the NIR information in all the relevant wavelengths into a small subset of artificial variables, called latent variables or regression factors $t_1, t_2, t_3, \dots, t_A$. Estimates of these factors $\mathbf{T} = (t_1, \dots, t_a, \dots, t_A)$ are used as regressors for y . They are also used as regressors for \mathbf{X} itself, in order to obtain residual spectra, which are useful as interpretation and diagnostic tools.

The estimated factors are often defined to be orthogonal to one another in the calibration set, since this simplifies the regression computations. The model for regressions on estimated latent variables can be summarized by

$$\mathbf{T} = w(\mathbf{X}) \quad (9.2)$$

$$\mathbf{X} = p(\mathbf{T}) + \mathbf{E} \quad (9.3)$$

$$y = q(\mathbf{T}) + f \quad (9.4)$$

where $w()$, $p()$, and $q()$ are mathematical functions to be defined in the following text, and \mathbf{E} and \mathbf{f} are residuals (representing measurement noise, model errors, etc.).

Notice that Equation (9.2) and Equation (9.4) can be rewritten and expressed by

$$\mathbf{y} = q(w(\mathbf{X})) + \mathbf{f} = b(\mathbf{X}) + \mathbf{f} \quad (9.5)$$

In the bilinear versions of these methods functions $w()$, $p()$, and $b()$ represent linear functions. The actual values of these have been estimated statistically, based on empirical data $(\mathbf{x}_i, \mathbf{y}_i)$ for a set of calibration samples $i = 1, 2, \dots, I$. As the name implies, PLSR employs the principal of least squares in this estimation. The term *partial* reflects the fact that the various parameters are estimated in different partial modeling steps rather than in one major parameter estimation process.

Before going into the details of PLSR modeling, let us consider some general aspects of bilinear modeling applicable to other bilinear methods as well (e.g., PCR).

Prior to the bilinear modeling the different X variables may be scaled to similar noise standard deviations, in order to stop high noise variability in some variables from hiding low, but informative variability in others.

Normally these scaled X and y variables are centered around their means $\bar{\mathbf{x}}$, \bar{y} prior to the calibration analysis. The main reason is that centering normally gives better linear approximation of possible nonlinear structures. Multivariate calibration, after all, concerns how to make good local approximations to the true but unknown nonlinear X - y relationship for a certain type of samples measured in a certain way.

The bilinear regression models then express the many input variables $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K)$ in terms of a few factors $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A)$, and these factors are also used for modeling \mathbf{y} . This is given by

$$\mathbf{X} = 1\bar{\mathbf{x}}^T + \mathbf{TP}^T + \mathbf{E}_A \quad (9.6)$$

$$\mathbf{y} = \bar{y} + \mathbf{Tq} + \mathbf{f}_A \quad (9.7)$$

Variables $\mathbf{T} = (t_{ia}, i = 1, 2, \dots, I; a = 1, 2, \dots, A)$ are called the factor scores, and define the level or “concentration” (positive or negative) of factor a in objects $i = 1, 2, \dots, I$. A score of, for example, $t_{ia} = 0$ represents a sample i having the average level of factor a . Parameters $\mathbf{P} = (\mathbf{p}_{ka}, k = 1, 2, \dots, K; a = 1, 2, \dots, A)$ and $\mathbf{q} = (q_a, a = 1, 2, \dots, A)$ are correspondingly called the X and y loadings for the factors. They define how each individual X and y variables are related to each underlying latent factors \mathbf{t}_a . For instance, a factor loading $q_a = 0$ represents a factor a with no relevance for \mathbf{y} , while $p_{ak} = 0$ represents an X variable k that is uncorrelated to factor a . Thus each vector \mathbf{p}_a is an abstract difference spectrum with a positive or negative value for each wavelength $k = 1, 2, \dots, K$. Matrices \mathbf{E}_A and \mathbf{f}_A in the bilinear model are the X and y residuals after A factors as shown in the following

$$\mathbf{E}_A = \mathbf{X} - 1\bar{\mathbf{x}}^T - \mathbf{t}_1\mathbf{p}_1^T - \dots - \mathbf{t}_A\mathbf{p}_A^T \quad (9.8)$$

$$\mathbf{f}_A = \mathbf{y} - \bar{y} - \mathbf{t}_1q_1 - \dots - \mathbf{t}_Aq_A \quad (9.9)$$

The term *bilinear model* stems from the fact that \mathbf{X} (and \mathbf{y}) are linear functions of both scores \mathbf{T} and loadings \mathbf{P} (and \mathbf{q}). In the bilinear model the product sum $\mathbf{TP}^T = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_A\mathbf{p}_A^T$ is intended to span the different systematic variabilities in the NIR spectra, varying chemical constituents (e.g., different organic solvents as hydrocarbons, alcohols, and ketones) and varying physical phenomena (light scattering, temperature effects, etc.).

In practice, the model parameters have to be estimated from empirical data. Since the regression is intended for later prediction of \mathbf{y} and \mathbf{X} , the factor scores \mathbf{T} are generally defined as functions of \mathbf{X} : $\mathbf{T} = w(\mathbf{X})$. The major difference between calibration methods is how \mathbf{T} is estimated. For

instance, in Fourier regression $w()$ is defined as a series of trigonometric functions; in PCR it is estimated as a series of eigenvector spectra for $(\mathbf{X} - 1\bar{\mathbf{x}}^T)^T(\mathbf{X} - 1\bar{\mathbf{x}}^T)$, etc. In PLSR $w()$ is defined as a sequence of \mathbf{X} vs. \mathbf{y} covariances.

9.4.1.2 PLS1 Regression

The PLSR estimation of the bilinear model for one y variable (PLS1) can be expressed in different ways. In the present orthogonalized version the bilinear factors $a = 1, 2, \dots, A$ are estimated and eliminated one at a time.

For each new factor a the PLSR scores are estimated as a linear combination of the X residuals after the previous $a - 1$ factors, given by

$$\hat{\mathbf{t}}_a = \hat{\mathbf{E}}_{a-1} \hat{\mathbf{w}}_a \quad (9.10)$$

where the so-called loading weight spectrum $\hat{\mathbf{W}}_a = (\hat{\mathbf{w}}_{ak}, k = 1, 2, \dots, K)^T$ in PLSR is defined as the normalized covariance matrix between the centered y variable and the residual variability remaining after the previous $a - 1$ factors, $\hat{\mathbf{E}}_{a-1}$, as shown in the following

$$\hat{\mathbf{w}}_a = g_a(\mathbf{y} - 1\bar{y})^T \hat{\mathbf{E}}_{a-1} \quad (9.11)$$

There g_a is a coefficient chosen so that the sum of squares of $\hat{\mathbf{w}}_a = 1.0$. These residuals after the previous $a - 1$ factors, $\hat{\mathbf{E}}_{a-1}$ are defined by

$$\hat{\mathbf{E}}_{a-1} = \mathbf{X} - 1\bar{\mathbf{x}} - \hat{\mathbf{t}}_1 \hat{\mathbf{p}}_1^T - \dots - \hat{\mathbf{t}}_{a-1} \hat{\mathbf{p}}_{a-1}^T \quad (9.12)$$

The corresponding loadings for each factor, \mathbf{p}_a and q_a , are simply estimated by univariate linear regression (projection) of the X and y variables on $\hat{\mathbf{t}}_a$. Thus, PLSR consists of extracting a few linear combinations of the many X variables, $\hat{\mathbf{t}}_a, a = 1, 2, \dots, A$. These estimated factors $\hat{\mathbf{t}}_a$ are used as regressor for both \mathbf{X} and \mathbf{y} . Each X combination $\hat{\mathbf{t}}_a$ is defined so as to maximize its covariance with \mathbf{y} .

This PLSR model has some nice orthogonality properties: The matrix of loading weights $\hat{\mathbf{W}} = (\hat{\mathbf{w}}_a, a = 1, 2, \dots, A)$ first of all is orthonormal, in other words, $\hat{\mathbf{W}}^T \hat{\mathbf{W}} = \mathbf{I}$ (the identity matrix). In addition, the present orthogonalized version of PLSR produces orthogonal scores ($\hat{\mathbf{T}}^T \hat{\mathbf{T}}$ is diagonal). The price to be paid for the latter is the double set of loadings, $\hat{\mathbf{W}}$ and $\hat{\mathbf{P}}$. The loadings $\hat{\mathbf{P}} = (\hat{\mathbf{p}}_a, a = 1, 2, \dots, A)$ are not orthogonal like $\hat{\mathbf{W}}$. But usually $\hat{\mathbf{w}}_a$ and $\hat{\mathbf{p}}_a$ are very similar, so it suffices to study one of them graphically.

More detail on PLSR is given, for example, in References 8–13 and Reference 14. PLSR in multivariate calibration is covered in detail in Reference 11.

9.4.1.3 Calibration and Prediction

Multivariate calibration in PLSR consists of estimating the model parameters, including the optimal number of factors to use in the mode, A . One factor is estimated at a time, $a = 1, 2, \dots, A$, in terms of its parameters, $\mathbf{w}_a, \mathbf{t}_a, \mathbf{p}_a, q_a, \mathbf{E}_a$, and \mathbf{f}_a , before the next factor is estimated.

The final calibration model thus consists of parameters, $\bar{\mathbf{x}}, \bar{y}, A, \hat{\mathbf{W}}, \hat{\mathbf{P}}$, and plus various summary statistics based on $\hat{\mathbf{T}}, \hat{\mathbf{E}}_A$, and $\hat{\mathbf{f}}_A$ to be used for subsequent outlier detection diagnostics. In addition, each calibration sample is characterized with respect to scores $\hat{\mathbf{t}}_a$ and residuals.

All these parameters are useful for interpretation of the calibration data, detecting outliers in the calibration set, etc. (Some details on how to determine the optimal number of factors and to assess the quality of the calibration are given in the following text.)

Then, once estimated, this calibration model can be used for prediction of y_i from \mathbf{x}_i in new objects $i = 1, 2, \dots, I$. This can be done in two ways: For every new object i with input data \mathbf{x}_i , the scores t_{ia} are obtained by $t_{ia} = \mathbf{e}_{i,a-1}^T \mathbf{w}_a$, where $\mathbf{e}_{i,a-1}^T$ is expressed by.

$$\mathbf{e}_{i,a-1} = \mathbf{x}_i^T - \bar{\mathbf{x}} - \sum_{m=1}^{a-1} \hat{t}_{im} \hat{\mathbf{p}}^T \quad (9.13)$$

and y_i is predicted by the following equation:

$$\hat{y}_i = \bar{y} + \sum_{m=1}^{a-1} \hat{t}_{im} \hat{q}_m \quad (9.14)$$

Alternatively, y_i can equivalently be predicted by

$$\hat{y} = b_0 + \sum_{k=1}^K \mathbf{x}_{ik} \hat{\mathbf{b}}_k \quad (9.15)$$

These estimated regression coefficients $\hat{\mathbf{b}} = (b_k, k = 1, 2, \dots, K)$ are defined as by

$$\hat{\mathbf{b}} = \hat{\mathbf{W}}(\hat{\mathbf{P}}^T \hat{\mathbf{W}})^{-1} \hat{\mathbf{q}}^T \quad (9.16)$$

$$b_0 = \bar{y}_0 - \mathbf{X}^T \hat{\mathbf{b}} \quad (9.17)$$

9.4.1.4 Validation and Assessment

During calibration it is important to determine the optimal number of PLS factors to retain in the calibration model; using too few factors can leave important NIR structure unmodeled, and using too many factors draws too much measurement noise from the X and y data into the calibration model. To guard against modeling nonsense, only factors with predictive ability are accepted in the calibration model. Cross-validation is used here. The full PLSR calibration modeling was thus repeated ten times, each time regarding a new set of about 10% of the calibration objects $i = 1, 2, \dots, I$ as a local, temporary test set. When finished, every calibration object has been treated as an “unknown” test object. The estimated predictive errors after a factors were thus obtained in the root mean square error of prediction (RMSEP). RMSEP is mathematical expressed as

$$\text{RMSEP} = \sqrt{\frac{1}{I} \sum_{i=1}^I (y_i - \hat{y}_i)^2} \quad (9.18)$$

The optimal number of factors A was determined as the number of factors after which RMSEP no longer decreased meaningfully.

Graphical inspection of loadings \mathbf{p}_a , $a = 1, 2, \dots, A$ was also used as an extra validation; factors that primarily reflect random noise are ignored in the final model.

The corresponding lack-of-fit standard deviation of each of the X variables, based on \mathbf{E}_a , is defined as

$$s(e_k) = \sqrt{\frac{\sum_{i=1}^I (x_{ik} - \bar{x}_k - \sum_{a=1}^A t_{ia} \hat{p}_{ak})^2}{df/k}} \quad (9.19)$$

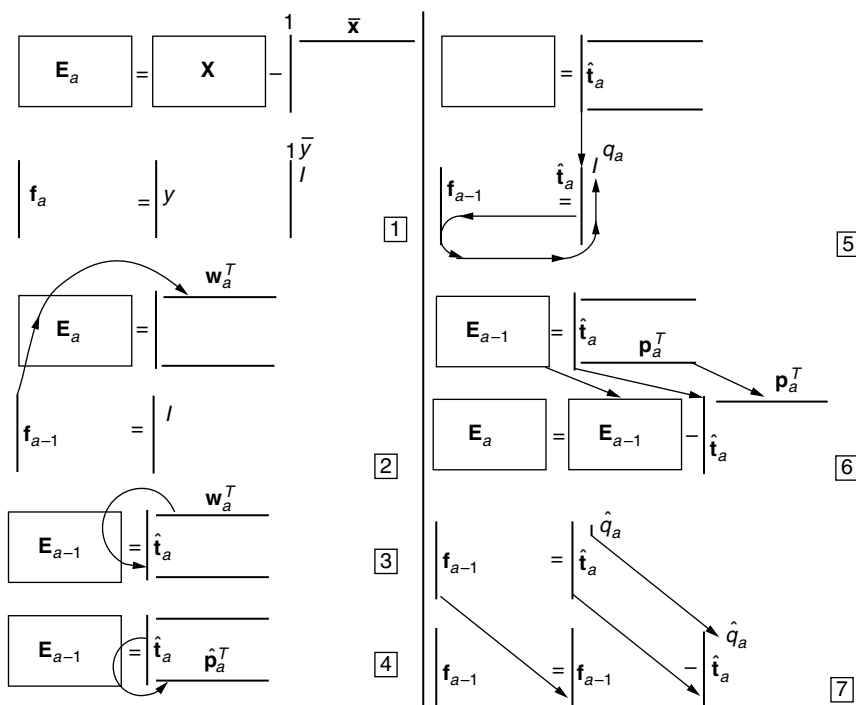


FIGURE 9.2 The PLS1 regression method is shown for modeling one y variable from KX variables in objects $i = 1, 2, \dots, I$. The figure shows that the variability not picked up by the A first bilinear factors remain in the residual matrices \mathbf{E} and \mathbf{f} . Each factor \mathbf{t}_a represents a linear combination of the X variables. Each latent variable also has a loading vector \mathbf{p}_a , which show how it related to the individual X variables, and another loading vector q_a that relates it to the y variable.

For the calibration set the degrees of freedom is here defined as $df = (IK - K - AK) = (I - 1 - A)K$. An analogous $s(e_k)$ for a test set is defined with $df = I(K - A)$. More detail is given in Reference 6 and Reference 7.

This two-block predictive PLS regression has been found very satisfactory for multivariate calibration and many other types of practical multivariate data analysis. This evaluation is based on a composite quality criterion that includes parsimony, interpretability, and flexibility of the data model; lack of unwarranted assumptions; wide range of applicability; good predictive ability in the mean square error sense; computational speed; good outlier warnings; and an intuitively appealing estimation principle. See, for example, Reference 6, Reference 7, and References 15–17.

The underlying notion in bilinear modeling is that “something causes the systematic variabilities in the X data.” But we may not correctly know what it is; there may be surprises in the data due to unexpected interferences, chemical interactions, nonlinear responses, etc. An approximate model of the subspace spanned by these “phenomena” in X is created. This X model is used for stabilizing the calibration modeling. The PLS regression primarily models the most dominant and most y -relevant of these X phenomena. Thus neither the manifest measured variables nor our causal assumptions about physical “laws” are taken for granted. Instead we tentatively look for systematic patterns in the data, and if they seem reasonable, we use them in the final calibration model.

By approximating complicated multivariate input data by a few bilinear factors, the user can plot simplified “maps” (in two or three dimensions of the main relevant information in the data). This allows the user — the person who needs the information and who knows the data and their context — to bring into use background knowledge and intuition in the interpretation, information that would be far too complex numerical information in the statistical modeling.

9.4.2 THE PLS1 ALGORITHM

In this section, the PLS1 method (PLSR for one y -variable) will be described in matrix algebra. The algorithm described here are the way the unscrambler program package derive a calibrating PLS1 model for predicting purposes (a model with only one y -variable). In Figure 9.2, the algorithm is illustrated schematically, where the arrows illustrate the way the data flow appear, and the boxes illustrates matrices, vectors, and scalar numbers. In Equation (9.20), the factor counter is initialized and set to zero. In Equation (9.21) and Equation (9.22) the data matrixes \mathbf{X} and \mathbf{y} , respectively, are centered. Steps 23 to 30 are repeated until the algorithm converges.

$$a = 0 \quad (9.20)$$

$$\mathbf{E}_a = \mathbf{X} - 1\bar{\mathbf{x}}^T \quad (9.21)$$

$$\mathbf{f}_a = \mathbf{y} - \bar{y} \quad (9.22)$$

$$a = a + 1 \quad (9.23)$$

$$\hat{\mathbf{w}}_{a^*}^T = \frac{\mathbf{f}_{a-1}^T \mathbf{E}_{a-1}}{\mathbf{f}_{a-1}^T \mathbf{f}_{a-1}} \quad (9.24)$$

$$\hat{\mathbf{w}}_a^T = \frac{\hat{\mathbf{w}}_{a^*}^T}{(\hat{\mathbf{w}}_{a^*}^T \hat{\mathbf{w}}_{a^*}^T)^{1/2}} \quad (9.25)$$

$$\hat{\mathbf{t}}_a = \mathbf{E}_{a-1} \hat{\mathbf{w}}_a^T \quad (9.26)$$

$$\hat{\mathbf{p}}_a^T = \frac{\hat{\mathbf{t}}_a^T \mathbf{E}_{a-1}}{\hat{\mathbf{t}}_a^T \hat{\mathbf{t}}_a} \quad (9.27)$$

$$\mathbf{E}_a = \mathbf{E}_{a-1} - \hat{\mathbf{t}}_a \hat{\mathbf{p}}_a^T \quad (9.28)$$

$$\hat{q}_a = \frac{\hat{\mathbf{t}}_a^T \mathbf{f}_{a-1}}{\hat{\mathbf{t}}_a^T \hat{\mathbf{t}}_a} \quad (9.29)$$

$$\mathbf{f}_a = \mathbf{f}_{a-1} - \hat{\mathbf{t}}_a \hat{q}_a \quad (9.30)$$

9.5 RESULTS AND DISCUSSION

Figure 9.1 showed that no single wavelength was selected for the analyte ethanol; the two interferents also absorb at every wavelength, although in somewhat different patterns.

The absorbance spectra $\log(1/T)$ were shifted to the same average baseline, based on the range between channels 41 (1305 nm) and 45 (1325 nm). These wavelength-corrected spectra represent the 101 X variables in the present dataset. The weight percentage of the analyte ethanol represents the y variable.

The X and y data for the 15 calibration mixtures were submitted to PLSR calibration.

9.5.1 RESIDUAL STATISTICS

Figure 9.3 shows how the average estimated predictive error (RMSEP) for the analyte ethanol (y) is reduced as the PLS1 model increases in complexity ($a = 0, 1, 2, \dots$). From chemical considerations one would expect two PLS factors to be sufficient in this set, since we have three constituents and their sum is 100%. The figure shows that most of the ethanol variability is correctly described by two PLS factors, but a minor third factor also seems to have predictive validity.

Figure 9.4 shows how the variability $s(e_k)$ in the X spectra of the calibration set is progressively described by the factors.

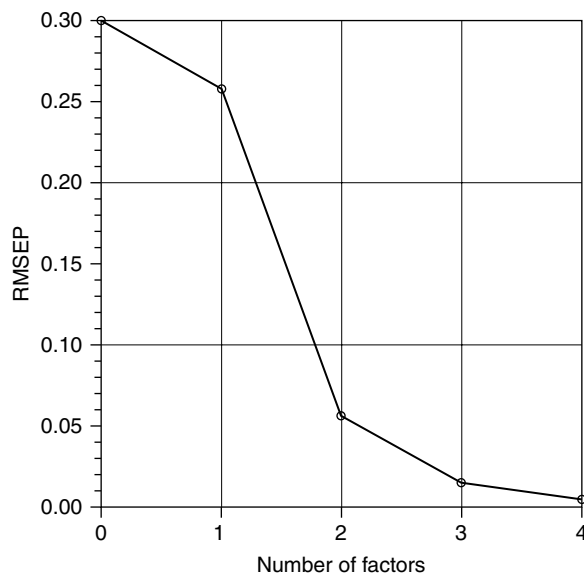


FIGURE 9.3 Estimated root mean square error of prediction for ethanol (y) RMSEP vs. the number of factors, $a = 1, 2, \dots$

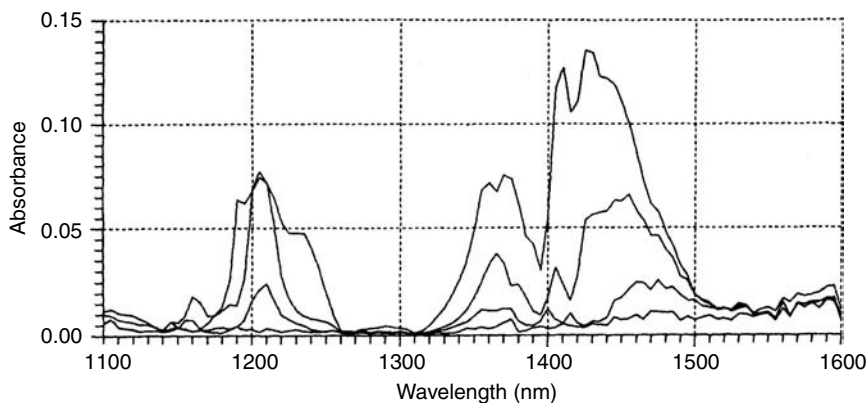


FIGURE 9.4 Estimated residual standard deviation, $s(e_k)$ vs. wavelength at X channels $k = 1, 2, \dots, K$ after $a = 0, 1, \dots, 3$ PLS factors.

9.5.2 PLSR MODEL

Figure 9.5 to Figure 9.9 show the resulting PLSR model for the X spectra in terms of the average $\bar{\mathbf{x}}$ (" $a = 0$ ") and the loading weights \mathbf{w}_a and loadings \mathbf{p}_a for factors $a = 1, 2, 3$. The figure shows that the first three factors seem to model smooth spectral phenomena, while the remaining factors primarily reflect noise. The two types of loading spectra are seen to be reasonably similar. Usually the individual PLS factors cannot be expected to reflect individual chemical or physical phenomena in the X data, like the spectral contributions by the analyte alone. For instance, in the present case the interferences' concentration will have to decrease when the analyte concentration increases. In general, the PLS factors will reflect combinations of different chemical and physical phenomena.

To study the actual chemical and physical phenomena affecting the X data, it can therefore be more informative to study combinations of PLS factors. Figure 9.10 shows a two-way plot of the

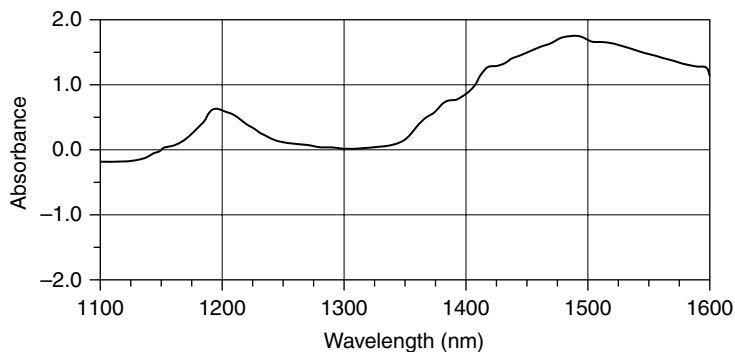


FIGURE 9.5 Model center $\bar{\mathbf{x}}$ corresponding to mean of $\bar{y} = 0.311$.

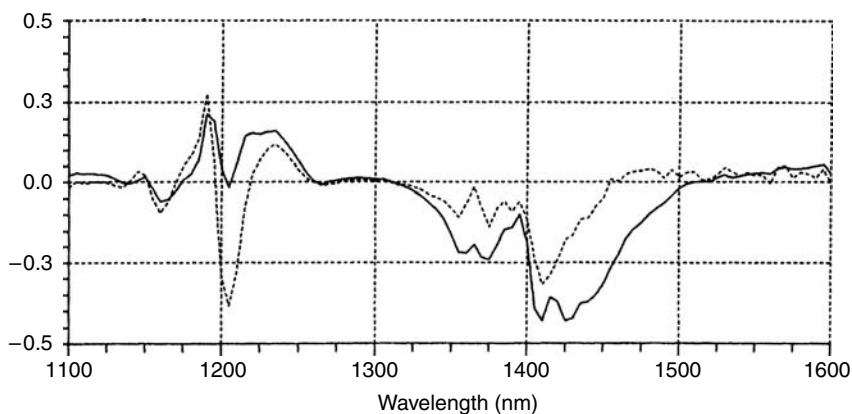


FIGURE 9.6 Loading weight $\hat{\mathbf{w}}_1$ and loading $\hat{\mathbf{p}}_1$, corresponding to $q_1 = 0.34$ and an explained sum of squares in \mathbf{X} , $\mathbf{t}_1^T \mathbf{t}_1 = 1.195$.

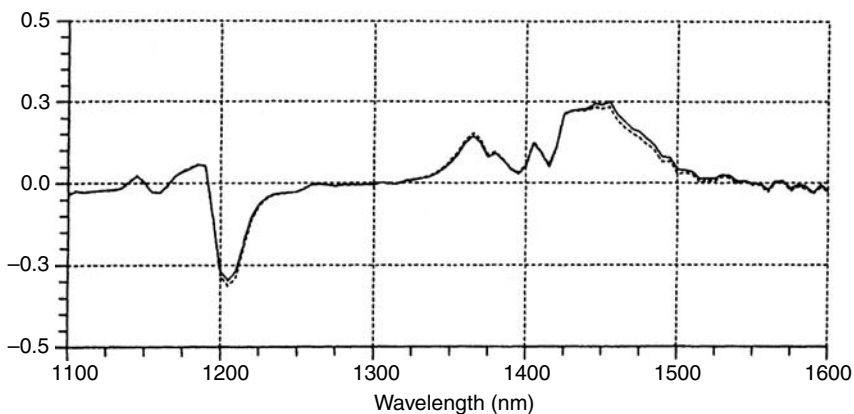


FIGURE 9.7 Loading weight $\hat{\mathbf{w}}_2$ and loading $\hat{\mathbf{p}}_2$, corresponding to $q_2 = 0.986$ and an explained sum of squares in \mathbf{X} , $\mathbf{t}_2^T \mathbf{t}_2 = 0.824$.

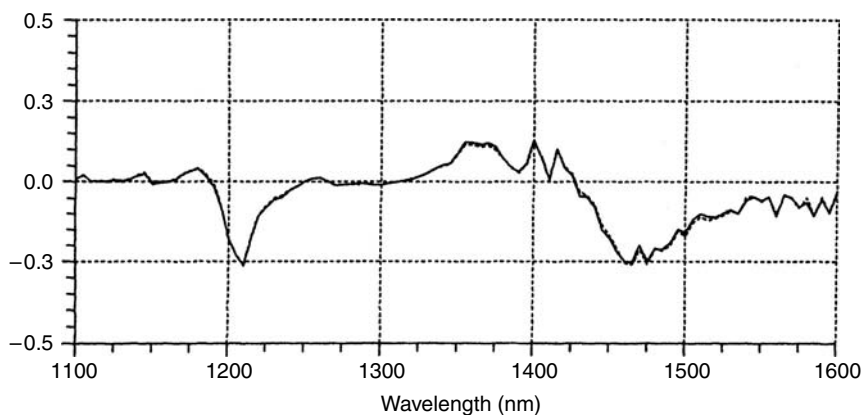


FIGURE 9.8 Loading weight \hat{w}_3 and loading \hat{p}_3 , corresponding to $q_3 = 0.341$ and an explained sum of squares in \mathbf{X} , $\mathbf{t}_3^T \mathbf{t}_3 = 0.103$.

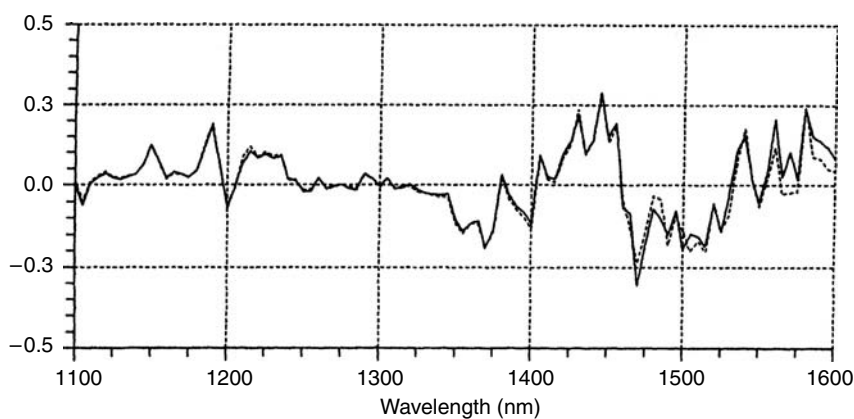


FIGURE 9.9 Loading weight \hat{w}_4 and loading \hat{p}_4 , corresponding to $q_4 = 0.472$ and an explained sum of squares in \mathbf{X} , $\mathbf{t}_4^T \mathbf{t}_4 = 0.005$.

loadings \mathbf{p}_a for factors $a = 1$ and $a = 2$. Each point in this plot represents a certain wavelength. Figure 9.11 shows the corresponding score plot for \mathbf{t}_a . Each point here represents an object in the calibration set. The positions of samples containing pure analyte and the two pure interferences methanol and *n*-propanol are marked in the plot. The scores reveal a triangular shape, corresponding to the ternary mixture design. But a slight nonlinearity is apparent; this was reflected in the third PLS factor, which also had predictive ability. The reason for this nonlinearity is unknown, but it arises primarily for mixtures of the two alcohols with the largest difference in refractive index and molecular weight. Similar effects have been observed using the same fiberoptics cuvette in other solvent mixtures like benzene, cyclohexane, and toluene. Figure 9.3 to Figure 9.9 indicate that the optimal number of PLS factors is $A = 3$. Figure 9.12 summarizes the corresponding PLS predictor in terms of the estimated regression coefficient $\hat{\mathbf{b}}$. It shows that the analyte ethanol is optimally determined as a weighted sum of the X variables with positive weights in wavelength regions 1130–1200, 1230–1260, and 1355–1365 and negative weights in regions 1155–1170, 1200–1230, and 1400–1440 nm.

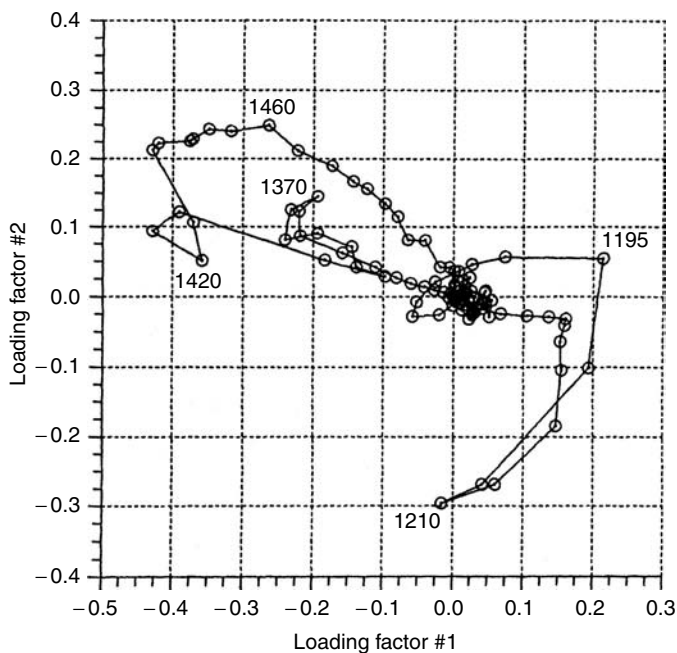


FIGURE 9.10 Two-way loading plot for the two first factors in the PLSR model. Abscissa: loading p_1 . Ordinate: loading p_2 . Each point represents a wavelength. Adjacent wavelength channels have been connected by straight lines; some variables have been marked explicitly for interpretation purposes.

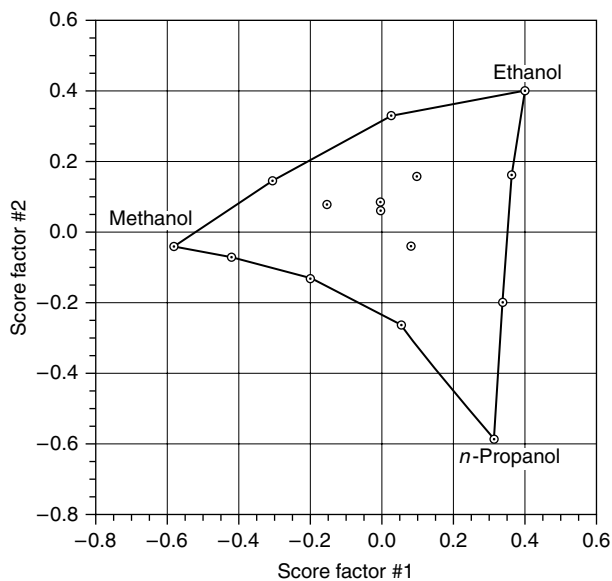


FIGURE 9.11 Two-way score plot for the first two factors in the PLSR model. Abscissa: score t_1 . Ordinate: t_2 . Each point represents an object. The binary mixtures in the design have been connected by straight lines. The corners of the triangle gives the positions of the analyte ethanol and the “unknown” interferents methanol and *n*-propanol.

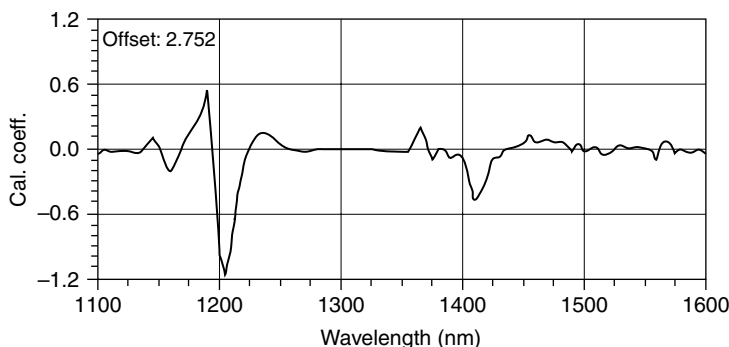


FIGURE 9.12 Calibration result for three PLS factors, summarized by the estimated regression coefficients \mathbf{b} plotted against wavelength. The corresponding offset $b_0 = 2.752$.

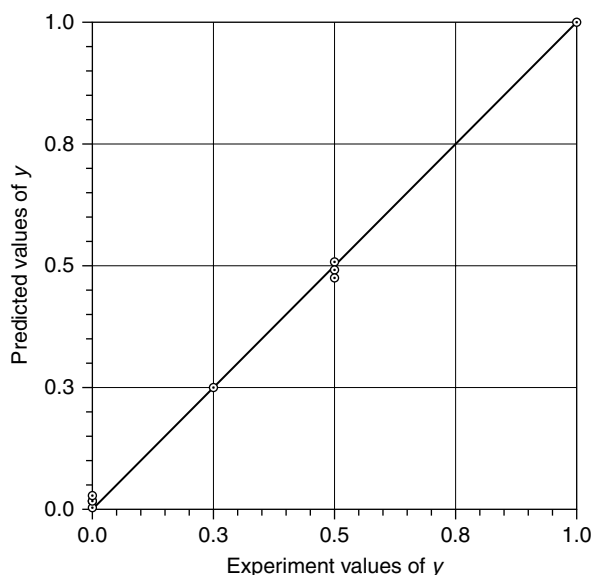


FIGURE 9.13 Predicted vs. known concentration of the analyte ethanol, of the 11 objects in the test set. Three PLS factors were used in the predictions.

9.5.3 PREDICTION OF ANALYTE IN NEW OBJECTS

The NIR spectra \mathbf{x}_i for the 11 test objects $i = 1, 2, \dots, 11$ was fitted to the obtained PLSR model, and the level of the analyte ethanol was predicted. The results are compared to the known analyte levels in Figure 9.13. A very good predictive ability ($R = .99$, RMSEP = 0.018 was obtained).

9.6 DISCUSSION

How “good” is PLSR? The performance of regression methods in NIR calibration can be assessed by theoretical considerations and by empirical comparisons on real datasets. The former is cheaper, while the latter may be more realistic. Both approaches are valuable when assessing PLSR. But both also have their problems.

The PLS1 regression seeks to optimize two things at the same time, spanning the main types of NIR variation (compressing \mathbf{X} into $\hat{\mathbf{T}}$) and relating the chemical data to the NIR data (regressing \mathbf{y}

on \hat{T}). Its performance hence depends on several things: the structure within the NIR data X and their noise levels, the X - y relationships, and the noise level in y . Therefore PLSR is more difficult to study theoretically than, for example, the traditional NIR or mixture models under standard statistical assumptions.

But NIR calibration is always a complex statistical systems analysis; irrespective of regression method, NIR calibration is usually done interactively. Background knowledge is used as part of the modeling method. So theoretical evaluations have to be used with caution for any NIR calibration method.

On the other hand, the predictive ability of a calibration model is very difficult to assess empirically. Even the computational methods that we use are questionable: Do squared errors correspond to the loss function that the user really wants? And the statistical requirements are important: How many test samples should be needed to test the predictive ability to what detail, and how should these test samples be distributed statistically in the multivariate sense? These are difficult questions for statisticians and chemists alike. On this basis, how can we assess PLSR as an NIR calibration method?

From theoretical considerations and from our experience, the following can be said about the performance of PLSR in NIR calibration: for high-quality calibration data (low noise and good linearity for X and y); PLSR will usually give the same prediction ability as many other multivariate calibration methods. The choice of method is then a matter of taste and habit.

Compared to PCR, PLSR usually gives more of the relevant modeling in the first few factors, making the graphical inspection of the model easier.

Compared to, say, SMLR the bilinear PLSR calibration model is usually easier to understand: The SMLR (and PLSR) regression coefficient vector \hat{b} is a contrast between all the different phenomena affecting the NIR spectra, so it can be very confusing. The bilinear PLSR allows us to study these NIR phenomena more or less individually, by graphical inspection of the A -dimensional model subspace.

If there is particularly grave nonlinearity in the X - y relationships, and a surplus of X variables is available, then SMLR may sometimes give slightly better predictive ability than PLSR and other full-spectrum methods, since the SMLR can simply skip the X variables that display the gravest nonlinearity for y , while the full-spectrum methods may have to model them.

On the other hand, if there is appreciable random noise in the NIR data (X), then full-spectrum methods must be expected to give better predictive precision than SMLR due to the full-spectrum smoothing effect.

If the chemical data (y) are noisy and the number of calibration samples low, then PLSR may or may not give an advantage over other methods. On the one hand, PLSR then overfits more easily than, for example, PCR, since the noisy y -variable is used more extensively in PLSR and in PCR. In that respect PLSR resembles SMLR, where y is used for both variable selection and parameter estimation. So the validation to determine the optimal number of factors is then very important in PLSR. On the other hand, we have observed that if the X data contain a lot of variability irrelevant for modeling y , then PLSR has a better chance, than, for example, PCR of extracting just the y -relevant X structures before over-fitting.

In summary, we have never found a case where PLSR has performed clearly worse than, say, SMLR or PCR, and we have found many cases where PLSR performs clearly better (in terms of understandable calibration models and low predictive errors). Thus we regard PLSR, with proper *a priori* weighting of the input variables, proper outlier detection, and conservative validation, to be a relatively safe and informative calibration method of general applicability.

However, it is essential to regard the design/calibration/interpretation/prediction/outlier detection process in NIR analysis as a part of the general cycle of scientific learning. In this context the choice of calibration method is primarily a question of what gives the best insight and the best outlier warnings in practice, not what gives the lowest statistical mean square error under idealized conditions. This means that the software implementation is just as important as the regression method itself.

ACKNOWLEDGMENTS

The authors express their gratitude to Randi Marie Hoem of Nerliens Kemisk Tekniske A. S. Oslo, and Arne Rosdahl (†) of Guided Wave International, Helsingborg, for making available the NIR spectrophotometer. The first author gratefully acknowledges Director of Chemistry R & D Ole Henrik Eriksen at Nycomed AS for the opportunity to do the research necessary for this chapter.

REFERENCES

1. K. G. Jöreskog and H. Wold, *Systems Under Indirect Observations, Causality/Structure/Prediction*, Vols. I and II, North-Holland, Amsterdam, 1983.
2. S. Wold, H. Martens, and H. Wold, The multivariate calibration problem in chemistry solved by PLS method. In: *Proc. Conf. Matrix Pencils* (A. Ruhe and B. Kågström, eds.), Lecture Notes in Mathematics, Springer-Verlag, Heidelberg, March 1982, pp. 286–293.
3. H. Martens and S. A. Jensen, Partial least squares regression: A new two-stage NIR calibration method. In: *Progress in Cereal Chemistry and Technology*, Vol. 5a (J. Holas and J. Kratochvil, eds.), Elsevier, Amsterdam, 1983, pp. 607–647.
4. Umscrambler Users Guide, Version 2.0. Software package for multivariate calibration and prediction. General version: CAMO A/S, Jarleveien 4, N-7041 Trondheim, Norway; Spectrometry version: Guided Wave Inc., 5200 Golden Foothill Pkwy, El Dorado Hills, CA 95630, 1987.
5. Grapher Reference Manual, Version 1.75. Golden Software, Inc., 807 14th Street, P. O. Box 281, Golden, Colorado 80402, 1988.
6. H. Martens and T. Næs, *Multivariate Calibration*. John Wiley & Sons Ltd., New York, 1989.
7. H. Martens and M. Martens, *Multivariate Analysis of Quality. An Introduction*. John Wiley & Sons Ltd., New York, 2001.
8. I. S. Helland, *Biometrics*, 43: 61–70 (1987).
9. A. Höskuldsson, PLS regression methods. *J. Chemometrics*, 2: 211–228 (1988).
10. P. Geladi and B. R. Kowalski, *Anal. Chem. Acta*, 185: 1–17 (1986).
11. H. Martens and T. Næs, *Multivariate Calibration*. John Wiley & Sons, New York.
12. T. Næs, C. Irgens, and H. Martens, *Appl. Stat.*, 195–206 (1986).
13. Wold et al. In: *Chemometrics: Mathematics and Statistics in Chemistry* (B. R. Kowalski, ed.), D. Reidel, Dordrecht, The Netherlands, 1984.
14. S. Wold, A. Ruhe, H. Wold, and W. J. Dun III, *Siam J. Sci. Stat. Comp.*, 5: 735–743 (1984).
15. H. Martens and T. Næs, Multivariate calibration by data compression. In: *Near-Infrared Technology in the Agricultural and Food Industries, 2nd edn.* (P. Williams and K. Norris, eds.), American Association of Cereal Chemists, St. Paul, 2001.
16. M. Martens (ed.), Data approximation by PLS methods. *Proceedings Symposium*, May 1987. Report No. 800 from the Norwegian Computing Center, P. O. Box 114, N-0314 Oslo 3, Norway.
17. M. Martens and H. Martens, Partial least squares regression. In: *Statistical Procedures in Food, Research* (J. R. Piggott, ed.), Elsevier, London, 1986, pp. 293–359.
18. H. Martens, Multivariate calibration, combining harmonies from an orchestra of instruments into reliable predictors of chemical composition, Invited paper on the 46th session of the ISI, Tokyo, 1988.

10 Aspects of Multivariate Calibration Applied to Near-Infrared Spectroscopy

Marc Kenneth Boysworth and Karl S. Booksh

CONTENTS

10.1	Introduction	207
10.2	Data Preprocessing	208
10.2.1	Mean Centering and Variance Scaling	208
10.2.2	Multiplicative Scatter Correction	210
10.2.3	Orthogonal Signal Correction	211
10.2.4	Instrument Standardization	213
10.3	Multivariate Calibration	214
10.3.1	Multiple Linear Regression	215
10.3.2	Principal Components Regression	215
10.3.3	Partial Least Squares Regression	216
10.3.4	Locally Weighted Regression	218
10.4	Model Selection and Validation	218
10.4.1	Outlier Detection	219
10.4.2	Model Selection	220
10.5	Further Analysis of Corn Flour Spectra	221
10.6	Optical Regression	224
10.6.1	Theory	225
10.6.2	Application	226
	Acknowledgments	228
	References	228

10.1 INTRODUCTION

Chemometrics is often defined as the application of statistics and mathematics to the analysis of chemical data. Without arguing the sufficiency of this definition, it is safe to say that the application of multivariate statistical and mathematical spectral analysis methods to near-infrared (NIR) data provides an intriguing set of advantages absent in univariate analysis of NIR data. Foremost of these advantages are the abilities to preprocess NIR spectra for removal of complex background signals, perform multianalyte calibration and calibration in the presence of multiple changing chemical

signals, and the ability to readily detect the presence of samples that deviate from the bulk of the calibration set.

This chapter consists of two distinct parts. In the first part, a cursory overview of chemometric methods, as applicable to analysis of and quantitation with NIR data is presented. In Section 10.2 common methods for preprocessing NIR spectra are described. Section 10.3 discusses multivariate methods for developing predictive calibration models with NIR spectra. In Section 10.4, strategies for sample and model validation are presented that exploit the multivariate nature of NIR spectra. The performance of the multivariate methods discussed in Section 10.2 through Section 10.4 are applied to a set of 80 NIR reflectance data of corn flour for determination of four physical properties: moisture, oil, protein, and starch.

The second part of this chapter presents a novel method for improving the precision of multivariate calibration with NIR spectra from scanning and filter-wheel spectrometers. *Optical regression* (OR) employs the regression model to determine the optimal operational parameters for the scanning monochromator or filter-wheel that maximizes the analytically useful signal-to-noise ratio. The theory of OR is presented in Section 10.6 and the method is applied to mixtures of dense nonaqueous phase liquids.

Naturally, it is impossible for one chapter to cover every aspect of chemometrics, as applicable to NIR spectral analysis, in sufficient detail. Instead, a representative sampling of chemometric methods is chosen and discussed. The goal is to present a range of accepted and commonly employed methods aimed at multivariate calibration with NIR data. These methods represent a basic foundation of chemometric analysis from which other new or more specialized methods evolve. Along with the chemometric methods of proven utility, a handful of new, promising analysis tools are interspersed in the discussion. These methods are chosen to be representative of the cutting edge of NIR analysis. While these newer methods are currently in the chemical literature, it is still too early to reliably predict their long-term impact.

10.2 DATA PREPROCESSING

The successful application of multivariate calibration methods is contingent upon the degree of uninformative variance in the NIR spectra. NIR spectra are commonly identified with low-intensity peaks occurring on the shoulders of more intense background spectra. These baselines may occur from black-body radiation from hot samples, scattering in turbid samples, or high concentrations of water in the sample. Concurrently, uninformative variance may be introduced into a data set by changes in the operational parameters of the NIR spectrometer. For example, the alignment of gratings, mirrors, and lamps, or employment of a different attenuated total reflectance cell could all alter the appearance of the NIR spectra.

This section describes promising multivariate methods for alleviating the effects of uninformative baseline variation. Other preprocessing methods, such as linearizing transformations of the spectra, smoothing functions, Fourier filters, derivatives, and estimation of a polynomial background are not discussed here. The absence of these popular methods is not meant as a critique of their efficacy; in many cases derivatives or smoothing are the ideal preprocessing methods for an application. However, these common methods are well documented in monographs dealing with univariate analyses.

10.2.1 MEAN CENTERING AND VARIANCE SCALING

Two of the most commonly employed methods of preprocessing multivariate data are *mean centering* and *variance scaling* of the spectra. Taken together, the application of mean centering and variance scaling is *autoscaling*.

Mean centering is applied by subtracting the mean spectrum of the data set from every spectra in the data set. For a data set $\mathbf{R}(I \times J)$ of I samples each of J discrete digitized wavelengths, the mean

centered j th wavelength of the i th sample is defined by

$$^{mc}R_{i,j} = R_{i,j} - \left(\sum_{j=1}^J R_{i,j} / J \right) \quad (10.1)$$

In a multivariate sense, this preprocessing method translates the collection of data to the origin of the multivariate space where analysis will be performed. The practical consequence of mean-centering data is often a more simple and interpretable regression model. In effect, mean centering removes the need for an intercept from the regression model. Consequently, since fewer terms in the regression model may need to be estimated, estimated analyte concentrations may be more precise following mean centering of the data. It should be noted that mean centering does not always yield the most precise calibration model. Each calibration method should be tested on mean centered and nonmean-centered data.

The effect of mean centering is demonstrated in Figure 10.1, Figure 10.2a, and Figure 10.2b. Figure 10.1a presents the raw spectra of the 40 corn flour samples employed in Section 10.5, while Figure 10.1b presents the mean-centered spectra. Although the spectra do not appear to be visually interpretable, none of the variance within the data set has been altered. The major effect of mean centering is removing the broad sloping background from the data collection. The effect of mean centering is shown in the cartoons of Figure 10.2a and Figure 10.2b. The data cloud in the upper right corner of Figure 10.2a is translated to the origin of the J dimensional space. The arrows of Figure 10.2 present the direction of greatest variance from the origin. For the nonmean-centered data, the direction of greatest variance is the mean of the spectra. With mean-centered data, the direction of greatest variance is now the direction of greatest variance *within* the data set. Consequently, more

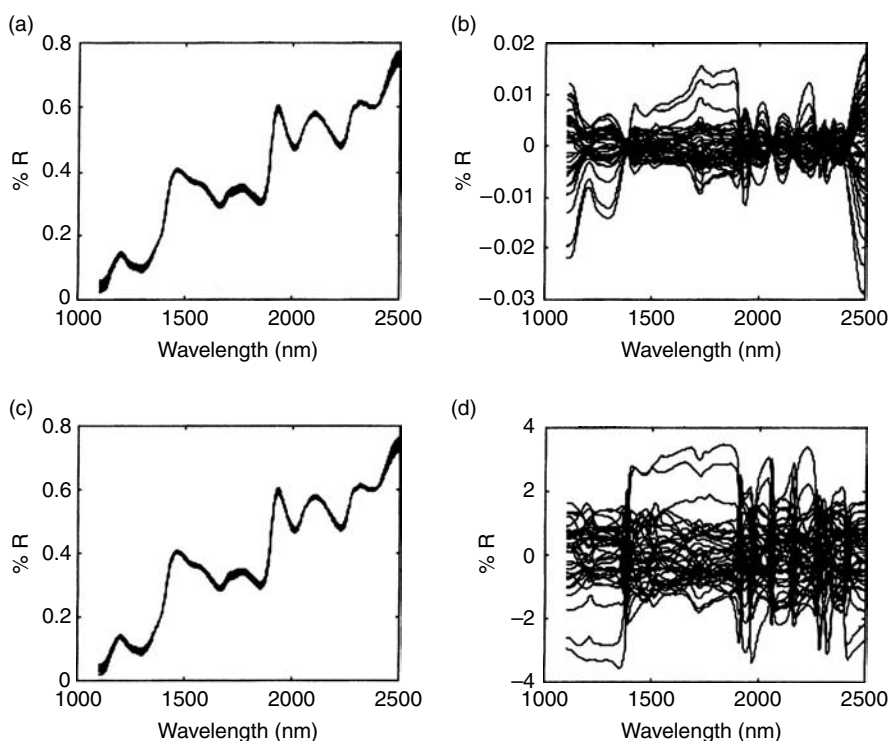


FIGURE 10.1 Forty corn calibration samples: (a) no pretreatment, (b) mean centered, (c) variance scaled, and (d) autoscaled.

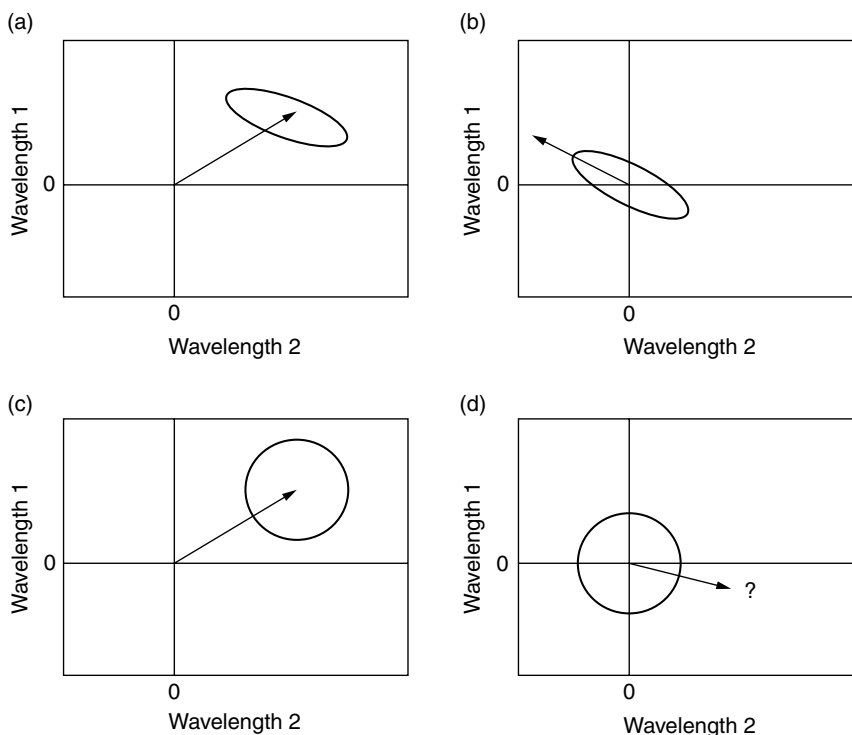


FIGURE 10.2 Cartoon representation of (a) untreated data, (b) mean-centered data, (c) variance-scaled data, and (d) autoscaled data. Arrows designate direction of greatest variance.

of the information content of a data set can usually be described with a more simple model if the data is mean centered.

10.2.2 MULTIPLICATIVE SCATTER CORRECTION

Multiplicative scatter correction (MSC) was developed to reduce the effect of scattered light on diffuse reflection and transmission NIR spectra [1,2]. This method has also shown utility as a means of removing varying background spectra with nonscattering origins. Consequently, MSC sometimes appears as multiplicative signal correction. The basic application of MSC is presented here. However, a more advanced version of MSC exists that assumes a unique scattering model for different regions of the spectra also exists [3].

Scattering theory states that scattering should have a multiplicative effect on reflection (and transmittance) spectra. That is, the observed spectra will contain a broad, changing background from differential scattering at each wavelength. In Figure 10.3a it is apparent that the largest source of variance within the 40 samples is derived from scattering. Assuming a multiplicative model for the scattering, the scattering profile in a spectrum can be deduced from a plot of the spectrum of a standard scatterer vs. a given spectrum at each wavelength. An ideal “standard” would have no NIR absorbance (or transmittance) features; however the mean spectrum from a collection of similar samples will suffice. Figure 10.3b presents the plot of the intensity of each wavelength for the mean of 40 calibration spectra vs. two of the individual calibration spectra. Note that one is scattering more than the average spectrum and one is scattering less than the average spectrum.

The plot for each of these two samples lies about a line with a little variation around the line. The difference between each sample and the best fit line through each sample in Figure 10.3b can

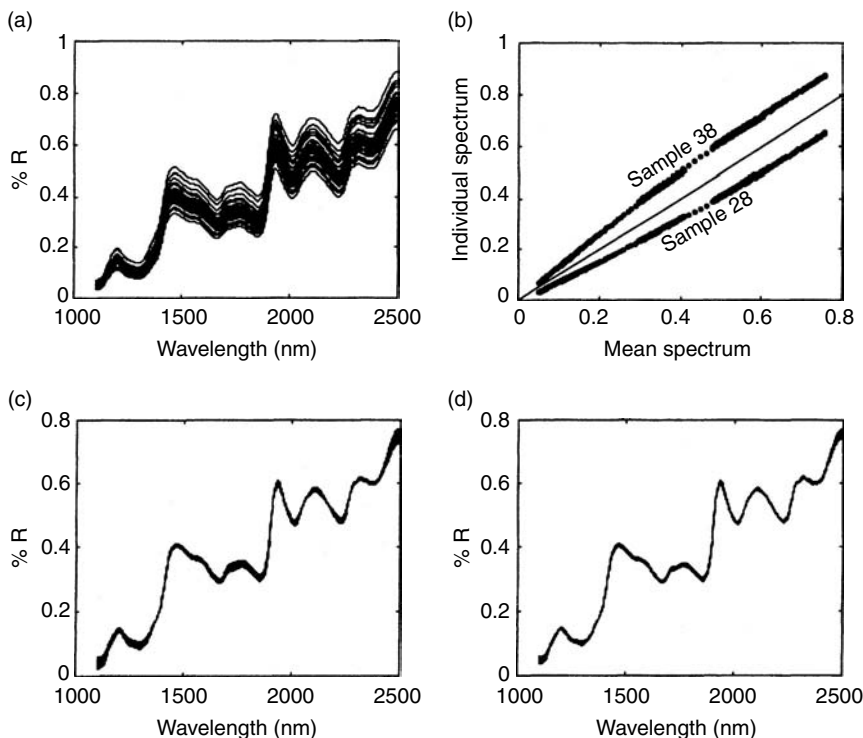


FIGURE 10.3 Forty corn calibration samples: (a) untreated, (b) plot of samples 38 and 28 vs. average of 40 samples, (c) MSC applied to 40 calibration samples, and (d) MSC applied to 40 future samples.

be interpreted as the chemical signal and the best fit line gives the spectrum of the scattering in the sample. Consequently, the scattering is determined by regressing each spectrum onto the mean spectrum, where the scattering at the j th wavelength of a sample can be modeled by

$$x_j = a + b\bar{x}_j + \varepsilon_j \quad (10.2)$$

with a and b being constant for all J wavelengths in the sample. The scatter-corrected data is determined by the scaled deviations about the regression:

$$X_{j,\text{MSC}} = (X_{j,\text{raw}} - a)/b \quad (10.3)$$

The corrected spectra for the 40 calibration NIR corn flour samples is shown in Figure 10.3c. For correction of future samples, the mean of the calibration set may be employed as the scatter standard. Figure 10.3d shows the corrected spectra of 20 corn flour spectra that were not included in the calibration set. Evident from these figures is that the spectral features are not distorted by MSC contrasted to scatter correction by calculating the second derivative of each spectrum.

10.2.3 ORTHOGONAL SIGNAL CORRECTION

Orthogonal signal correction (OSC) was recently developed as an alternative to MSC for NIR spectra. OSC is fueled by the realization that the majority of the spectral variance in an NIR data set is of little or no analytical value. Therefore, variance that is orthogonal to the property of interest is removed from the data set. Three algorithms have been published. The original iterative algorithm employs alternating orthogonalization and regression steps to determine the orthogonal signals [4].

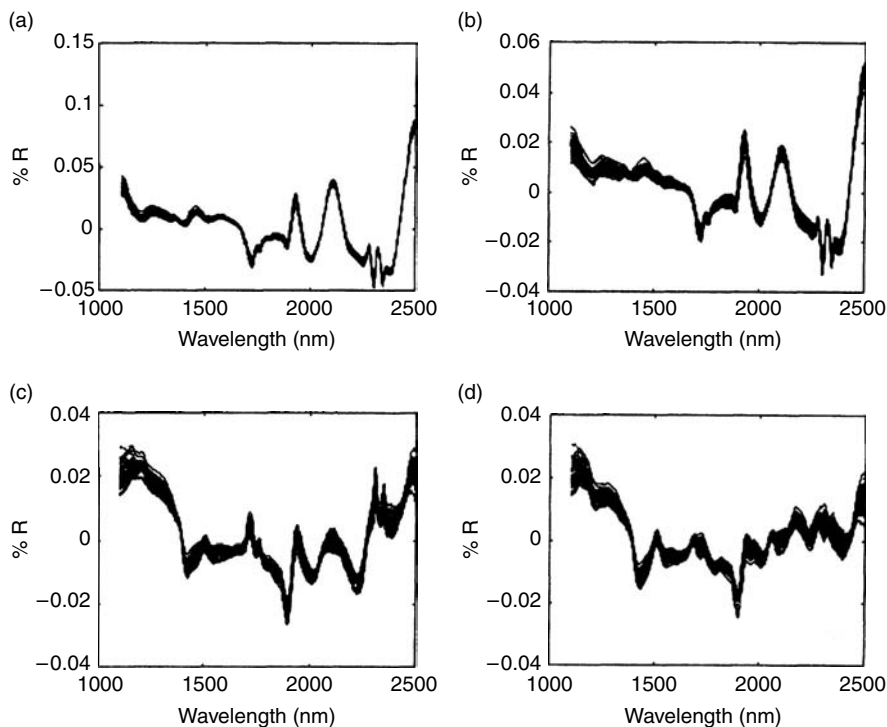


FIGURE 10.4 OSC applied to corn flour spectra: (a) one-factor correction for moisture, (b) two-factor correction for moisture, (c) three-factor correction for moisture, and (d) four-factor correction for moisture.

A second algorithm iteratively orthogonalizes the data prior to a regression step to determine the orthogonal instrument signal [5]. The final algorithm employs an eigenproblem to directly determine the orthogonal instrumental signal [6]. Future spectra are corrected by projecting each future spectra into the OSC model.

Unlike MSC, OSC requires a degree of optimization to determine the best scatter-correction protocol for a given application. The ideal number of orthogonal “factors” to be removed from the data set must be determined; the more factors eliminated, the greater the reduction of orthogonal variance. Figure 10.4a–d presents the OSC-corrected spectra of 40 corn flour calibration spectra for determination of moisture. Figure 10.4a presents OSC with a one-factor correction, Figure 10.4b presents OSC with a two-factor correction, and Figure 10.4c and Figure 10.4d present OSC with a three- and four-factor correction, respectively. Two shortcomings of OSC, compared to MSC, are evident. First, the corrected spectra do not look like NIR spectra. This is a consequence of part of the “real” NIR signal being discarded with part of the background scattering. Therefore, it is difficult to achieve an intuitive, visual, feel for the sufficiency of the signal correction. Second, as more orthogonal signal is discarded, the remaining spectral intensity decreases. This is logical because both the orthogonal scattering and orthogonal NIR spectra are removed from the data set.

A significant difference between OSC and MSC is highlighted in Figure 10.5. Figure 10.5a through Figure 10.5b present the two-factor correction of the same 40 corn flour samples for determination of oil, moisture, protein, and starch, respectively. On the one hand, OSC requires a unique scatter-correction model to be determined for each analyte in a data set. Also, the success of the signal correction will be partially dependent on the accuracy of the analyte to be calibrated (i.e., moisture content). Conversely, only one MSC model needs to be determined for a range of analytes and the value of the analytes do not need to be known to determine the correction model. On the other hand, OSC calculates a scatter-correction model that is optimized for a particular application. This may, in fact, lead to an improved calibration model relative to MSC corrected data.

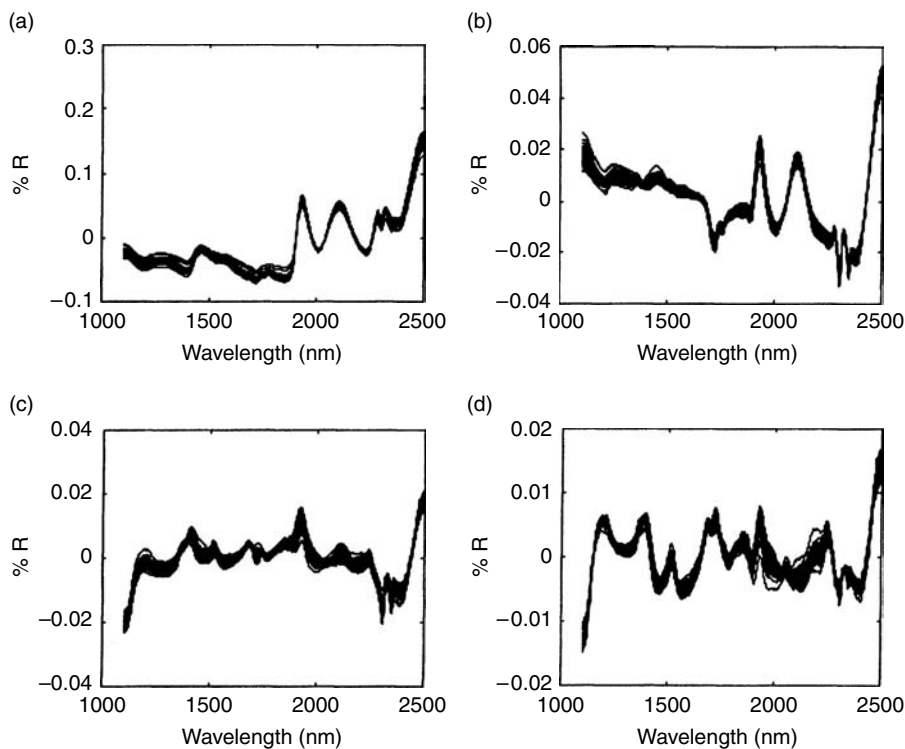


FIGURE 10.5 OSC applied to corn flour spectra: (a) two-factor correction for oil, (b) two-factor correction for moisture, (c) two-factor correction for protein, and (d) two-factor correction for starch.

10.2.4 INSTRUMENT STANDARDIZATION

MSC and OSC address changes in the spectral backgrounds, or baselines, that occur from changes in the sample matrix. However, spectral changes often occur from “drift” in the spectrometer. To mitigate the effects of spectrometer drift, or to better permit simultaneous analysis of spectra from different spectrometer, *instrument standardization* is employed. The goal of standardization is to develop a transformation function that makes spectra collected at one point in time appear as if they were collected under the same instrumental conditions as a previous set of “standard” spectra. The ideal implementation of instrument standardization would have a high-resolution, laboratory spectrometer be the standard instrument; spectra collected on low-resolution, inexpensive spectrometers would be standardized to become higher resolution spectra.

One standardization method popular in the literature is piecewise direct standardization (PDS) [7–10]. With PDS, a set of transfer samples are analyzed on both the original instrument and the instrument to which the calibration model will be transferred. It is best if the transfer samples are a subset of the calibration set; however, other surrogate samples may be employed. A separate transfer function is determined for each wavelength in the spectra by least squares regression using neighboring wavelengths as the independent variables. That is, a local subset of variables measured on the second instrument is employed to build a model that predicts what each measurement would have been if it were measured with the first instrument. This method accounts for shifts and intensity changes over a small spectral window. The drawback of PDS is that success of the standardization is dependent on choice of the transfer samples. The transfer samples must be identical when measured on each instrument and the set of samples must span the space of all encountered spectral changes between the two instruments. Therefore, the choice and number of transfer samples must be optimized by the analyst.

A more useful method of standardization would not require transfer samples to be analyzed. A general method based loosely on MSC has also demonstrated success when there are relatively minor performance differences between the original and second instruments [11,12]. Here a local selection of wavelengths from each spectrum are regressed against the mean spectrum to build a transfer function. Consequently, the spectra from the second instrument are not transformed to look like the spectra from the first instrument. Instead, spectral from both instruments are transformed to lie in a common multidimensional space.

10.3 MULTIVARIATE CALIBRATION

The goal of calibration is to relate the observed spectra, in a predictive manner, to a property of interest. In the example presented here, NIR spectra are employed to moisture, oil, protein, and starch in corn flour samples. Multivariate analysis has three intrinsic capabilities that make a powerful tool for calibration applications [13]. First, complete signal selectivity is not required for successful implementation of multivariate calibration methods. With univariate calibration, a NIR signal must be determined that is unique to the analyte of interest, in other words, protein. Obviously, in a complex sample such as corn flour, finding a unique wavelength for an analyte is impossible. Second, multivariate analysis allows for calibration of a multicomponent sample. Not only can a calibration model be constructed for more than one analyte, but the model can be constructed for the analyte in the presence of multiple spectral interferents. And finally, multivariate analysis permits powerful diagnostic tools to aid in assessing the quality and reliability of future spectra. Such diagnostics do not exist for univariate calibration methods.

There are two paradigms of multivariate calibration. Classical least squares (CLS) models the instrumental response as a function of analyte concentration.

$$\mathbf{R} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (10.4)$$

where each row of \mathbf{R} is the spectrum of a sample, \mathbf{C} is a matrix (samples \times constituents) of the extrinsic property of interest (often concentration) and \mathbf{S} are NIR spectra of each constituent in \mathbf{R} . Equation (10.4) follows a logical Beer's law-type model where the instrument response is expressed as a function of the constituent's concentration and spectra.

For a set of recorded spectra, \mathbf{R} , with known constituent extrinsic properties, \mathbf{C} ; \mathbf{S} can be estimated by ordinary least squares inversion of \mathbf{C}

$$\hat{\mathbf{S}}^T = \mathbf{C}^+ \mathbf{R} \quad (10.5)$$

Premultiplication of a future recorded spectrum, \mathbf{r}_{un} , by the (pseudo)inverse of \mathbf{S} affects the estimation of extrinsic properties in the unknown sample,

$$\hat{\mathbf{c}}_{\text{un}} = \hat{\mathbf{S}}^+ \mathbf{r}_{\text{un}} \quad (10.6)$$

Consequently, there are advantages and disadvantages associated with CLS. The primary advantage is that estimates of the true constituent spectra are derived during calibration. However, CLS requires knowledge of the concentrations of all constituents with a spectral profile in the calibration set. From a practical standpoint, this may be an unreasonable requirement. In many cases, in other words, spectroscopic estimation of octane number or Reid vapor pressure, no "spectrum" of the extrinsic property would exist.

The alternative to the CLS calibration model is the inverse least squares (ILS) calibration model. Employing an ILS model alleviates the need for complete knowledge of the calibration set

constitution. With the ILS model, concentration (or any intrinsic property) is modeled as a function of instrument response

$$\mathbf{c} = \mathbf{R}\mathbf{b} \quad (10.7)$$

where the vector \mathbf{b} relates the variance in the observed data, \mathbf{R} , to \mathbf{c} . The *regression vector* \mathbf{b} is chosen to be correlated to \mathbf{c} and contravariate (orthogonal) to all sources of variance in \mathbf{R} that are not correlated to \mathbf{c} . This regression vector is purely a mathematical construct and has no physical meaning. By employing such *soft-modeling* calibration based on correlation, knowledge of only the extrinsic property of interest to be estimated is required for calibration. Numerous regression methods based on the ILS model exist. These methods differ only in the nature of the regression vector or the manner in which the regression vector is calculated. Common linear ILS calibration methods are discussed in Section 10.3.1 to Section 10.3.4.

10.3.1 MULTIPLE LINEAR REGRESSION

For the linear calibration methods, the regression vector, \mathbf{b} , is estimated directly from \mathbf{R} and \mathbf{c} where

$$\hat{\mathbf{b}} = \mathbf{R}^+ \mathbf{c} \quad (10.8)$$

with the overscript $\hat{}$ representing an estimate and the superscript $+$ representing an arbitrary inverse of a matrix. The linear ILS methods differ in the determination of \mathbf{R}^+ . With multiple linear regression (MLR) \mathbf{R}^+ is determined from the normal equations,

$$\mathbf{R}^+ = (\mathbf{R}^T \mathbf{R})^{-1} \mathbf{R}^T \quad (10.9)$$

While this is perhaps the most straightforward methods of performing ILS calibration, MLR suffers from numerous shortcomings. First, collinearity in \mathbf{R} makes determination of the true inverse of $(\mathbf{R}^T \mathbf{R})$ an ill-condition problem. Collinearity in \mathbf{R} occurs when any column of \mathbf{R} can be expressed as linear combinations of the other columns of \mathbf{R} , or when there are more columns in \mathbf{R} than there are rows in \mathbf{R} . With NIR spectroscopy, the correlation of absorbances (or transmissions) of adjacent wavelengths leads to collinearity in \mathbf{R} . Also, there are usually fewer calibration samples available than there are recorded wavelengths in NIR spectra. Consequently, MLR often leads to unstable estimates of \mathbf{b} . The MLR model will fit the calibration set well. If the regression vector is unstable, small random errors in future samples will be magnified. This may result in large prediction error in future samples.

However, MLR should not be summarily rejected. For applications with a small number of wavelengths (i.e., data from filter-wheel spectrometers) MLR is potentially the most applicable ILS method. In some cases, application of MLR following judicious selection of a wavelength may outperform other ILS calibration methods.

10.3.2 PRINCIPAL COMPONENTS REGRESSION

One common and robust ILS calibration method is principal components regression (PCR) [14,15]. The first step of PCR is factorization of \mathbf{R} with the principal component analysis (PCA) decomposition where

$$\mathbf{R} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (10.10)$$

The orthonormal columns of \mathbf{U} define the sample space of \mathbf{R} and the orthonormal columns of \mathbf{V} define the wavelength space of \mathbf{R} . The columns of \mathbf{U} and \mathbf{V} are arranged in decreasing order of variance of \mathbf{R} described. The scale of \mathbf{U} and \mathbf{V} lies in the diagonal matrix \mathbf{S} . Mathematically the columns of \mathbf{U} are the eigenvectors of $\mathbf{R}^T \mathbf{R}$, the columns of \mathbf{V} are the eigenvectors of $\mathbf{R}\mathbf{R}^T$ and the diagonal elements of \mathbf{S}^2 are the eigenvalues of either matrix. Some factors of \mathbf{R} (columns of \mathbf{U} and \mathbf{V}) are

related to systematic chemical information; other factors mostly describe random instrumental and measurement errors. Under the assumption that the NIR spectral variations from sample to sample are greater than the instrumental noise, the first few factors that describe the chemical signal are deemed significant. The remaining factors are discarded for calibration.

Once significant factors are determined, error reduction is accomplished by discarding factors that describe random variance. Thus

$$\mathbf{R} \approx \overline{\mathbf{U}}\mathbf{S}\overline{\mathbf{V}}^T = \overline{\mathbf{R}} \quad (10.11)$$

where the overscript implies only the significant factors are retained. Because PCR (and PLS as follows) discards variance from the calibration (truncated columns of \mathbf{U} and \mathbf{V}) set prior to regression, these methods are said to be biased. However, any loss of accuracy is compensated by an improvement in precision. The net result is a general lessening of prediction errors compared to employing all of the variance of \mathbf{R} in calibration as with MLR.

The second step in PCR is the estimation of the regression vector \mathbf{b} . Either a subset of the scores, $\overline{\mathbf{U}}$, or the random-error reduced response matrix, $\overline{\mathbf{R}}$, can be related to the dependent variable, \mathbf{c} . However, both options are equivalent. The orthonormal character of the scores and loadings make estimation of \mathbf{b} trivial once \mathbf{R} has been factorized by Equation (10.10). Because both $\overline{\mathbf{U}}^T\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}^T\overline{\mathbf{V}}$ are identity matrices,

$$\hat{\mathbf{b}} = \overline{\mathbf{R}}^+ \mathbf{c} = \overline{\mathbf{V}}\mathbf{S}^{-1}\overline{\mathbf{U}}^T \mathbf{c} \quad (10.12)$$

where the inverse of the square diagonal matrix \mathbf{S} is the inverse of the elements along the diagonal. Estimation of \mathbf{c} for future samples is accomplished by multiplication of the regression vector with the recorded instrumental response of the sample in question,

$$\hat{\mathbf{c}}_{\text{un}} = \mathbf{r}_{\text{un}}^T \hat{\mathbf{b}} \quad (10.13)$$

Table 10.1 presents the spectral and concentration variance described for successively more complex (more factors) PCR calibration models for oil in corn flour. The data are processed by MSC and mean centering the spectra and concentrations. Each successive factor describes less spectra variance. The first factor contains 87% of the total spectral variance about the mean of the data set while the second and third factors contain 6 and 2% respectively. The remaining factors describe less than 1% of the spectral variance each. However, while the first factor does describe most of spectral information, the first factor only accounts for 3% of variation in oil content. With the most important factors for predicting oil concentration being the second (47%), eighth (21%), and fourth (10%) factors, eight factors are required to describe 90% of the oil concentration variation. Therefore, factors with a low signal-to-noise ratio have a large effect on the accuracy of the calibration model.

10.3.3 PARTIAL LEAST SQUARES REGRESSION

Partial least squares regression [16–18] has been employed since the early 1980s and is closely related to PCR and MLR [18]. In fact, PLS can be viewed as a compromise midway between PCR and MLR [19]. In determining the decomposition of \mathbf{R} (and consequently removing unwanted random variance), PCR is not influenced by knowledge of the estimated property in the calibration set, \mathbf{c} . Only the variance in \mathbf{R} is employed to determine the latent variables. Conversely, MLR does not factor \mathbf{R} prior to regression; all variance correlated to \mathbf{c} is employed for estimation. PLS determines each latent variable to simultaneously optimize variance described in \mathbf{R} and correlation with \mathbf{c} , ρ . Technically, PLS latent variables are not *principal components*. The PLS factors are rotations of the PCA PCs for a slightly different optimization criterion.

TABLE 10.1
Spectral and Predictive Variance Captured by Successive PCR
Factors for Determination of Oil in Corn Flour Samples. The NIR
Reflection Spectra Were Pretreated by MSC and Mean Centering

Factors	Percent spectral variance	Cumulative % spectral variance	Percent predictive variance	Cumulative % predictive variance
1	88.21	88.21	3.53	3.53
2	6.57	94.78	47.07	50.60
3	2.16	96.94	0.01	50.61
4	0.98	97.92	10.26	60.86
5	0.76	98.68	2.07	62.93
6	0.46	99.14	8.83	71.76
7	0.30	99.45	0.33	72.10
8	0.25	99.70	21.06	93.15
9	0.11	99.81	0.01	93.16
10	0.06	99.87	0.67	93.83
11	0.03	99.90	0.30	94.14
12	0.03	99.93	0.64	94.78
13	0.02	99.95	1.10	95.88
14	0.01	99.96	0.08	95.97
15	0.01	99.97	0.40	96.37

Mathematically the first PLS factor is determined to maximize

$$(\text{VAR} \times \rho)^2 = \mathbf{s}^4 \frac{(\mathbf{u}_1^T \mathbf{c})^2}{\mathbf{c}^T \mathbf{c}} \quad (10.14)$$

Successive PLS factors maximize Equation (10.14) under the constrained orthogonality to prior PLS factors. The PLS factors are determined by iterative algorithms that are described in most chemometric texts [20–22] and in references [15,18]. In fact, numerous algorithms exist that are optimized for various sizes of \mathbf{R} [23,24].

PLS has two distinct advantages compared to PCR. First, PLS generally provides a more parsimonious model than PCR. PCR calculates factors in decreasing order of \mathbf{R} -variance described. Consequently, the first factors calculated, that have the least imbedded errors, are not necessarily most useful for calibration. On the other hand, the first few PLS factors are generally most correlated to concentration. As a result, PLS achieves comparable calibration accuracy with fewer latent factors in the calibration model. This further results in improved calibration precision because the first factors are less prone to imbedded errors than are lower variance factors.

Second, the PLS algorithm is often faster to implement and optimize for a given application than is the PCR algorithm. PLS calculates the factors one at a time. Hence, only the latent variables needed for calibration are determined. PCR, employing the singular value decomposition, calculates all possible principal components for \mathbf{R} prior to regression. For large data sets that require relatively few factors for calibration, PLS can be significantly faster than PCR.

Compared to PCR, PLS leads to more parsimonious calibration models. Where PCR required eight factors to describe 90% of the oil variation, PLS only requires four factors (Table 10.2). By determining the factors such that spectral and predictive variance are simultaneously maximized, PLS can achieve the 90% predictive level while employing only 94% of the spectral variation. PCR required 99.7% of the spectral variance.

TABLE 10.2
Spectral and Predictive Variance Captured by Successive PLS Factors
for Determination of Oil in Corn Flour Samples. The NIR Reflection
Spectra Were Pretreated by MSC and Mean Centering

Factors	Percent spectral variance	Cumulative % spectral variance	Percent predictive variance	Cumulative % predictive variance
1	82.54	82.54	13.96	13.96
2	12.19	94.73	43.35	57.30
3	0.84	95.58	26.28	83.58
4	0.57	96.15	9.15	92.73
5	1.78	97.94	0.84	93.58
6	0.78	98.72	0.83	94.40
7	0.61	99.33	0.44	94.84
8	0.11	99.44	1.65	96.49
9	0.22	99.66	0.34	96.83
10	0.12	99.78	0.51	97.34
11	0.07	99.85	0.42	97.77
12	0.05	99.90	0.49	98.25
13	0.02	99.92	0.45	98.70
14	0.01	99.93	0.41	99.11
15	0.02	99.96	0.20	99.31

10.3.4 LOCALLY WEIGHTED REGRESSION

The global linear models calculated by PCR are PLS are not always the best strategy for calibration. Global models span the variance of all the samples in the calibration set. If the data are nonlinear, then the linear PCR and PLS methods do not efficiently model the data. One option is to employ nonlinear calibration methods. The second option is to employ linear calibration methods on small subsets of the data.

The locally weighted regression (LWR) philosophy assumes that the data can be efficiently modeled over a short span with linear methods. The first step in LWR is to determine the N samples that are most similar with the unknown sample to be analyzed. Similarity can be defined by distance between samples in the spectral space [25] by projections into the principal component space [26] and by employing estimates of the property of interest [27]. Once the N nearest standards are determined, either PLS or PCR is employed to calculate the calibration model.

LWR has the advantage of often employing a much simpler, and more accurate, model for estimation of a particular sample. However, there are three disadvantages associated with LWR. First, two parameters must be optimized for LWR (number of local samples and number of factors) compared to just one parameter (number of factors) for PLS and PCR. Second, a new calibration model must be determined for every new sample analyzed. Third, LWR often requires more samples than PCR or PLS in order to build meaningful, local calibration models.

10.4 MODEL SELECTION AND VALIDATION

There are two important validations that may be performed during multivariate calibration. The first validation is to determine if any of the samples in the calibration set are *outliers*. Outliers may be samples with unique spectral features that do not occur in the bulk of the other samples, samples with extreme properties that are outside the range of the other samples, or calibration samples that have

been mislabeled. The validation test is determination and assurance of the optimal preprocessing and optimal number of factors in the calibration model.

10.4.1 OUTLIER DETECTION

Two important statistics for identifying outliers in the calibration set are the *sample leverage* and the *studentized residuals*. A plot of leverage vs. studentized residuals makes a powerful tool for identifying outliers and assigning probable cause. The sample leverage is a measure of the influence, or weight, each sample has in determining the parameters of the calibration model. Samples near the center of the calibration set (average samples) will have a relatively low leverage compared to samples at the extreme edges of the experimental design and outliers. The sample leverage is determined by

$$h_i = 1/I + \bar{\mathbf{u}}_i^T \bar{\mathbf{u}}_i \quad (10.15)$$

where $\bar{\mathbf{u}}_i$ is the row of \mathbf{U} associated with the significant factors for the i th sample. Consequently, the same leverage ranges from 0 for a sample in the center of an infinitely large calibration set to one for an extreme sample in a small data set.

The studentized residual is an indication of how well the calibration model estimates the analyte property in each sample. The studentized residual is similar to the Student's t -statistic; the estimation error of each sample is converted to a distance in standard deviations away from zero. An additional term is often added to the calculation to correct for the weight each sample has in determining the calibration model. The studentized residual is increased for samples with a large leverage; this is known as the *studentized leverage corrected residuals*.

The studentized leverage corrected residuals are calculated by

$$t_i = \frac{|c_i - \hat{c}_i|}{\sigma \sqrt{1 - h_i}} \quad (10.16)$$

where

$$\sigma = \sqrt{\frac{\sum_{i=1}^I (c_i - \bar{c})^2}{I - N - 1}} \quad (10.17)$$

with N being the number of factors in the calibration model.

The plot of studentized leverage corrected residuals vs. sample leverage provides insights into the quality of each calibration sample. Samples with low leverages and low studentized residuals are typical samples in the calibration set. Samples with large studentized residuals and small leverages are suspect for concentration errors (mislabeled samples). Samples with large studentized residuals and large leverages may have either unique spectra features or concentration errors. Large sample leverages and small studentized residuals may be caused by spectral errors or by perfectly good samples that happen to lie at the extremes of the experimental design. Therefore, these samples should be examined closely before they are removed from the calibration set.

Example plots of leverage vs. studentized residual are presented in Figure 10.6a through Figure 10.6d. The figures were constructed for a five-factor PLS calibration model with mean-centered, MSC-corrected spectra for determination of moisture, oil, protein, and starch, respectively. It has not yet been determined whether a five-factor PLS model is the best model for calibration. However a five-factor model is close enough to optimal to determine whether or not samples are obviously bad samples. In all four plots the majority of the samples have low leverages and low studentized residuals. The large studentized residuals of sample 39 in Figure 10.6b and samples 26 and 6 in Figure 10.6d draws scrutiny upon these samples. The recorder oil estimate of sample 39 and protein estimate of sample 36 are unremarkable. The values lie near the center of the range of analyte values. It would be impossible to speculate on the source of the prediction errors without knowledge of how the reference analysis was performed. However, the protein concentration of sample 6 is

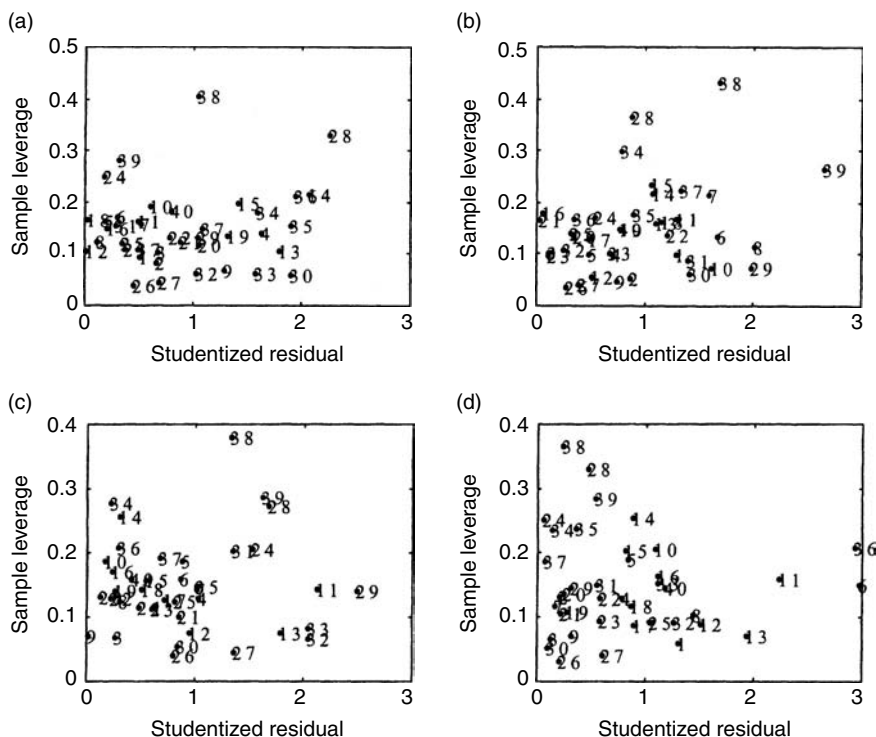


FIGURE 10.6 Outlier diagnostics for five-factor PLS model for calibration of (a) moisture, (b) oil, (c) protein, and (d) starch.

noteworthy. The difference in concentration between sample 6 and the closest protein concentration in the calibration set accounts for 20% of the spread of protein concentrations. Therefore, sample 6 could either be the recipient of a large experimental error or be in a nonlinear concentration regime.

Sample 38, and to a lesser extent sample 28, present high leverages for all four analyses. However, the studentized residuals for the two samples are not that large. Since these two samples have high leverages and low residuals for all four analyses, it is expected that they are more likely to contain spectral errors than to be extreme concentration values. Interestingly, sample 38 and sample 28 are the extreme samples for degree of scattering. As the greatest and least scattering samples. Sample 38 and sample 28 were chosen for Figure 10.3b. This implies that the MSC might not be perfect for correction of these spectra. Indeed, a slight dog-leg in Figure 10.3b for sample 38 suggests that a piece-wise MSC algorithm might be more appropriate.

10.4.2 MODEL SELECTION

Once the preprocessing and sample validation has been performed, the next tack is the optimization and validation of the calibration model. The goal here is to choose the most accurate and precise calibration model possible and to estimate how well it will perform in future samples. If a sufficient quantity of calibration samples is available, the best method for selecting and validating a model is to divide the calibration set into three subsets. One set is employed to construct all of the models to be considered. The second set is employed to choose the best model in terms of accuracy and precision. The third set is employed to estimate the performance of the chosen model on future data. Alternately, the data set can be divided into to subsets with the optimal calibration model being chosen by cross validation [28].

There are three statistics often employed for comparing the performances of multivariate calibration models: root mean squared error of calibration (RMSEC), root mean squared error of cross validation (RMSECV), and root mean squared error of prediction (RMSEP). All three methods are based on the calculated root mean squared error (RMSE)

$$\text{RMSE} = \left(\sum_{i=1}^I (c_i - \hat{c}_i)^2 / I \right)^{1/2} \quad (10.18)$$

where RMSEC, RMSECV, and RMSEP differ in the determination of \hat{c} .

The best estimate of future performance of a calibration model is the RMSEP. Concentration estimates, \hat{c} , in the RMSEP are determined by applying the calibration model to a subset of data that was not employed in determining the model parameters. The RMSEP may be calculated for a *validation set* in order to determine the optimal number of factors in a model or to a *test set* in order to test the performance of the optimal model on future data. If an external subset of data is not available to optimize the calibration model, the RMSEP can be estimated by the RMSECV. The concentration estimates of Equation (10.18) are determined in CV by iteratively removing (and replacing) each sample from the data set. When a sample is removed from the data set, a calibration model is constructed from the remaining samples. The property of the removed sample is then estimated by the calibration model. RMSEC is a measure of how well the calibration model fits the calibration set. This is potentially the least informative of the three statistics. RMSEC is an extremely optimistic estimation of the model performance. As more factors are included in the calibration model, the RMSEC always decreases.

The performances of the three statistics are evident in Figure 10.7 and Figure 10.8. Figure 10.7a–d presents the RMSEC, RMSECV, and RMSEP vs. number of factors for PCR calibration of moisture, oil, protein, and starch, respectively. All spectra were preprocessed by MSC and mean centered. Figure 10.8 is identical to Figure 10.7, except PLS was performed for calibration. The optimal number of factors estimated by statistical tests applied to the RMSE, choosing the first minima in the plot, or choosing the global minimum in the plot [28]. In this, and following analysis, the optimal model will be defined from the first minimum unless later minima yield “significantly” lower RMSE.

Note that the RMSEC is always decreasing along with the number of factors. In contrast, the RMSECV and RMSEP occasionally increase when more factors are included. As more factors are included in the calibration model, the model begins to fit the random errors imbedded in the spectra and concentrations. Therefore, the RMSEC will always decrease as more factors are added. However, new samples not included in the calibration set will have a different realization of random errors. Therefore, the calibration model will not fit these errors to the same degree as the errors in the calibration set. When extra factors that mostly describe random errors are included in the calibration model, these factors will not fit the errors in future samples and the RMSECV and RMSEP may increase.

Also note that the RMSEC usually provides the most optimistic estimate of model performance. In the limit, if every factor were included in the calibration model, the RMSEC would be zero. This is due to the ability of the calibration model to model the random errors in the calibration set; eventually every source of variation is modeled. However, the random errors in future samples cannot be precisely modeled. This is evident in the RMSEP and RMSECV yielding similar RMSE. Therefore, RMSECV is a better estimate of the future performance of model prediction than is RMSEC.

10.5 FURTHER ANALYSIS OF CORN FLOUR SPECTRA

Table 10.3 presents the RMSE for determination of moisture, oil, protein, and starch with PCR and PLS calibration. The 80 samples data set was divided into three subjects. Forty samples, selected

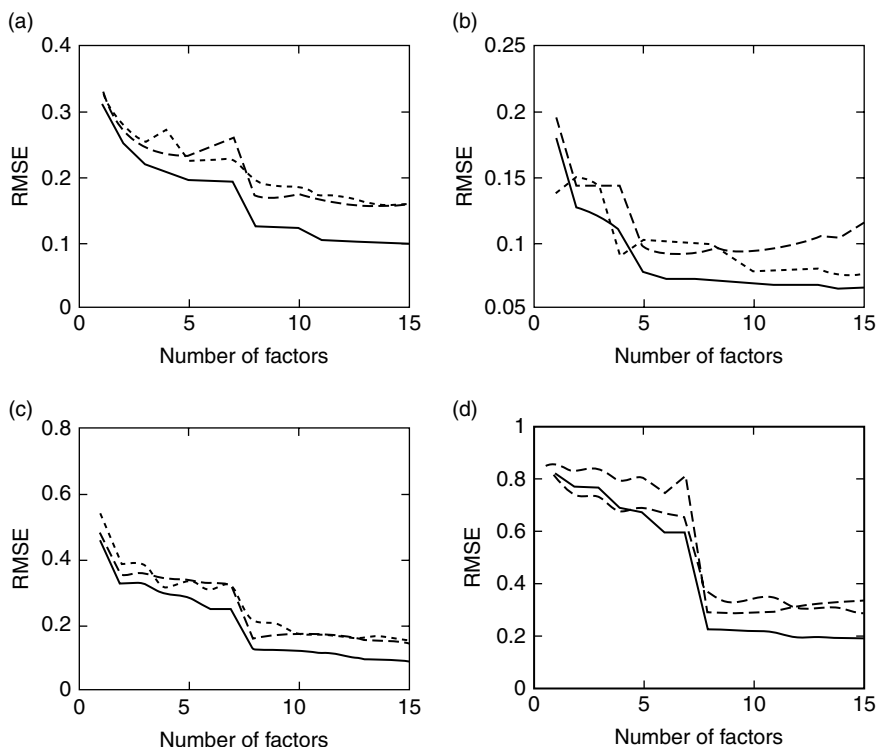


FIGURE 10.7 Model optimization statistics for PCR calibration of (a) moisture, (b) oil, (c) protein, and (d) starch. Solid lines indicated RMSEC, dashed line represents RMSECV, and dotted line represents RMSEP for the validation set.

randomly, were designated the calibration set. These samples were employed to calculate the RMSEC and RMSECV. The remaining 40 samples were split into two 20 sample groups: the validation and test sets. The 4th and 6th columns of Table 10.3 contain the optimal RMSEP calculated with the validation and test sets, respectively. The fifth and seventh columns contain the RMSEP for the test and validation sets as determined by the number of factors recommended by the other set. For example, for determination of moisture using PCR (first row). The RMSEP for the validation set recommends 11 factors be included in the model (column 4). When an 11-factor model is applied to the test set, a RMSEP of 0.167 is achieved. The eighth column of Table 10.3 presents the ensemble RMSEP for the validation and test sets as calculated from the number of factors recommended by cross validation. The predictive ability of the models is better visualized in Figure 10.9 where the true vs. estimated properties of the four analytes are plotted for the calibration, validation, and test sets with PLS calibration. The equivalent plots for PCR are qualitatively similar. In these plots the estimated concentration for the calibration set is presented for the number of factors recommended by cross validation. The estimated concentrations for the test and validation sets employed the models recommended by the RMSEP of each set.

Many of the trends discussed in the first four sections are evident in Table 10.3. PLS generally yields a more parsimonious calibration model than does PCR. PLS requires five to eight factors while PCR is recommending eight to ten factors. The RMSEC is the most optimistic estimation of model performance. However, the RMSEC usually recommends approximately the same number of factors as the RMSECV and the RMSEP (at least for the test set).

A comment on estimating the proper number of factors is warranted here. Note that there is usually a spread of recommended model complexities for each analyte. The optimal number of

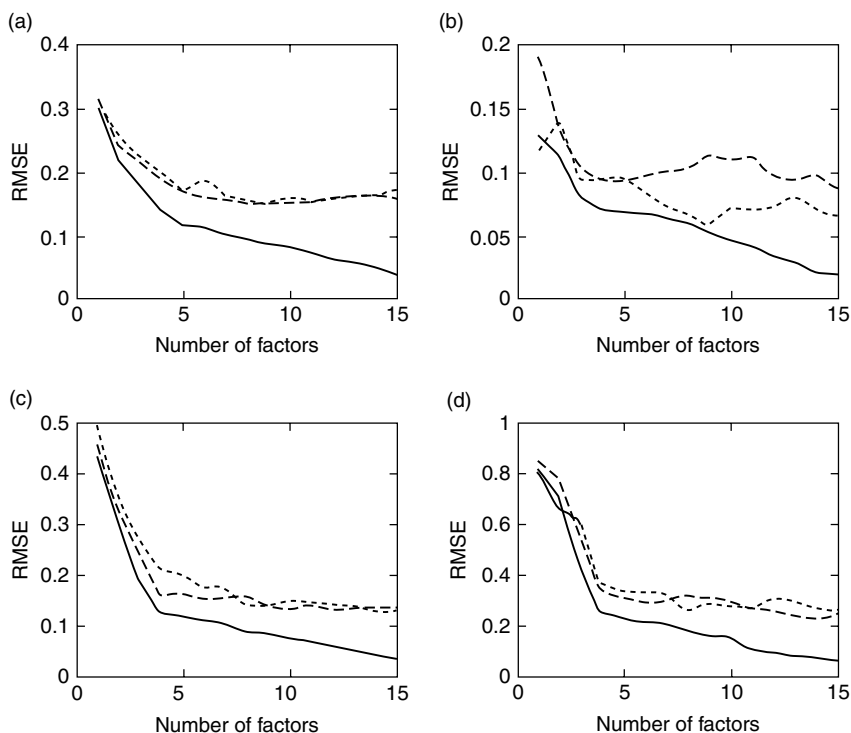


FIGURE 10.8 Model optimization statistics for PLS calibration of (a) moisture, (b) oil, (c) protein, and (d) starch. Solid lines indicated RMSEC, dashed line represents RMSECV, and dotted line represents RMSEP for the validation set.

TABLE 10.3

Root Mean Squared Errors for PCR and PLS-based Calibration of Corn Flour Samples. Parenthetical Numbers Indicate the Number of Factors Employed in the Calibration Model

	RMSEC ^a	RMSECV ^b	PMSEV ^c	$V \geq T^d$	RMSET ^e	$T \geq V^f$	$CV \geq V \& T^g$
PCR moisture	0.13 (8)	0.17 (8)	0.17 (11)	0.17	0.14 (8)	0.20	0.17
PLS moisture	0.12 (5)	0.15 (8)	0.15 (9)	0.19	0.13 (5)	0.17	0.17
PCR oil	0.07 (6)	0.09 (6)	0.08 (10)	0.10	0.09 (5)	0.08	0.10
PLS oil	0.07 (5)	0.09 (4)	0.06 (9)	0.10	0.08 (5)	0.11	0.09
PCR protein	0.12 (8)	0.15 (8)	0.17 (10)	0.17	0.14 (10)	0.14	0.18
PLS protein	0.09 (8)	0.16 (4)	0.14 (8)	0.20	0.15 (5)	0.15	0.19
PCR starch	0.23 (8)	0.30 (8)	0.33 (9)	0.36	0.24 (8)	0.32	0.35
PLS starch	0.25 (4)	0.29 (6)	0.26 (8)	0.33	0.30 (6)	0.27	0.32

^a Optimal root mean squared error of calibration.

^b Optimal root mean squared error of cross validation calculated from the calibration set.

^c Optimal root mean squared error of prediction calculated from the validation set.

^d Root mean squared error of prediction calculated from the test set with the model optimized from the validation set.

^e Optimal root mean squared error of prediction calculated from the test set.

^f Root mean squared error of prediction calculated from the validation set with the model optimized from the test set.

^g Root mean squared error of prediction calculated from the test and validation sets with the model optimized from cross validation of the calibration set.

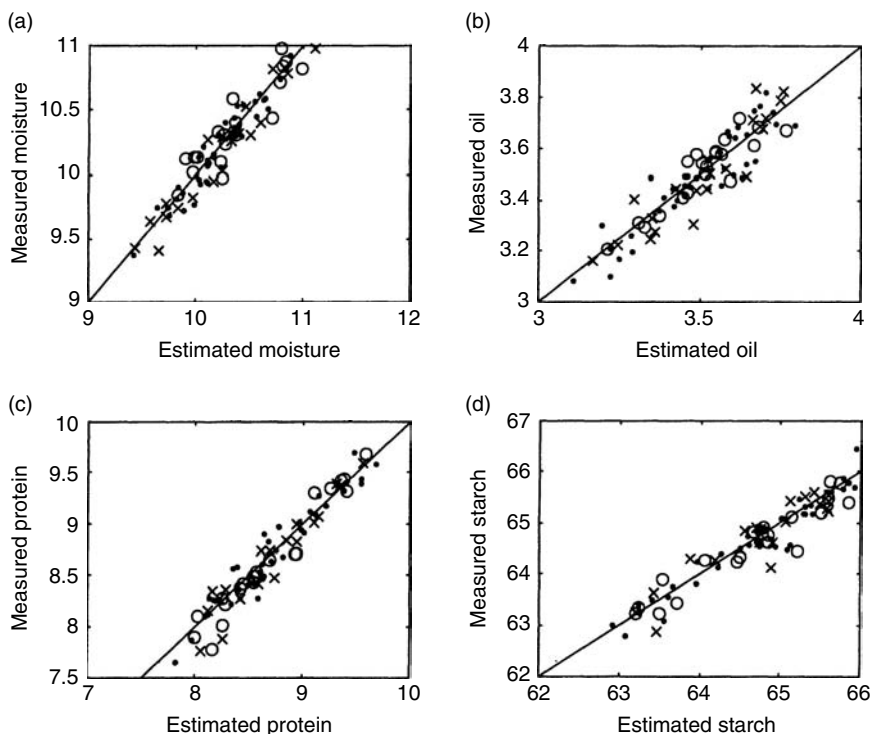


FIGURE 10.9 Plots of measured vs. estimated sample properties for (a) moisture, (b) oil, (c) protein, and (d) starch. Calibration samples are indicated by (●), test set samples by (x), and validation set samples by (○).

factors chosen depends on the data set employed to determine the best model. Similarly, there is a spread of estimated model performances. In some cases, the validation set presents lower RMSEP than the test set. Reasons for these two observations can be found in Figure 10.9. In Figure 10.9a, it is seen that the moisture values of the test set (x) do not span the same range as the calibration (●) or validation (○) sets. Likewise, for oil, the validation set is bunched in the middle of the range of oil concentrations while the test samples lie more at the extremes. This means that the different sets give more weight to different areas of the calibration set. Some regions of the experimental design might be harder to model than others, or some regions might be more prone to overfitting (employing random errors for correlation to concentration variance). Therefore, no single estimate of optimal model complexity of model performance should be considered an exact value. Each estimate should have error bounds assigned. However, the proper manner to assign the error bounds is not known!

10.6 OPTICAL REGRESSION

In the first five sections, the multivariate, NIR spectral preprocessing methods and determination of calibration models were presented. Multivariate regression models were calculated, optimized, and applied to digitized spectra of corn flour. Following determination of the multivariate model, the model will be applied to future spectra for estimation of the property of interest. With dispersive spectrometers, all wavelengths of the future spectra are collected simultaneously. The spectra are digitized, multiplied by the regression vector and the analyte concentration estimate is achieved (Equation (10.13)). However, with scanning and filter-wheel spectrometers, uniform digitization of spectra prior to computer manipulation is not necessarily the optimal method for applying multivariate

estimation. This section presents an alternative method of applying a multivariate regression model with scanning and filter-wheel spectrometers.

OR is designed to optimize the analytical precision of estimation for multivariate analysis with scanning or filter-wheel spectrometers. That is, the standard deviation of \hat{c} in Equation (10.13) will be minimized. OR employs the regression vector as a template for optimal collection of the spectral images [29]. In essence, the magnitude in the regression vector at each wavelength relays the relative importance of that wavelength for the overall analysis. Wavelengths associated with small values in the regression vector contain relatively little analytical information; large regression-vector values denote analytically significant wavelengths. Consequently, the analytically important signal-to-noise ratio can be maximized by spending more time integrating wavelengths with large regression-vector values at the expense of time devoted to unimportant measurements.

OR can be viewed in two contexts. On one level, it is a method for dynamic development of optimal experimental designs. Once a measurement problem is defined — that is, the regression vector is calculated — OR determines the optimal manner to collect future spectra that maximize predictive precision under the constraint of constant analysis time. This can also be inverted to minimize analysis time while maintaining constant accuracy and precision. Alternately, OR is a step toward an integrated optical computer. Much of the data analysis is performed on the detector — alleviating the need of many digitization steps. Consequently, errors associated with digitization and reading the detector are minimized. Performing more of the computations on the detector also minimizes the postcollection, computational time, and requirements.

10.6.1 THEORY

OR is derived from statistical propagation of error theory. Consider a set of J wavelengths, \mathbf{r} , where the observed errors are independently and identically distributed along a Gaussian distribution with a mean of zero and a standard deviation of σ . Given a regression vector, \mathbf{b} , the magnitude of the projection \mathbf{r} onto \mathbf{b} is indicative of a sample quality or quantity, \hat{c} . This quantity parameter is hence estimated by Equation (10.13) as

$$\hat{c} = (\mathbf{r} + \mathbf{e})^T \mathbf{b} = \sum_{j=1}^J (r_j + e_j) b_j \quad (10.19)$$

where the errors in \hat{c} are

$$e_{\hat{c}} = \mathbf{e}^T \mathbf{b} = \sum_{j=1}^J e_j b_j \quad (10.20)$$

The expected standard deviation of these errors in prediction are given as,

$$\sigma_{\hat{c}} = \left(\sum_{j=1}^J \sigma_j^2 b_j^2 \right)^{1/2} = \sigma \left(\sum_{j=1}^J b_j^2 \right)^{1/2} = \sigma \|\mathbf{b}\|_2 \quad (10.21)$$

where $\|\mathbf{b}\|_2$ is the two-norm, or euclidian norm, of \mathbf{b} . Equation (10.21) gives the standard deviation of prediction for traditional multivariate regression where each of the J wavelengths are equivalently recorded — that is, the same integration time is dedicated to all J measurements. This is digital regression (DR).

Alternately, the relative integration time for each measurement can be adjusted, under the constraint of constant total integration time, to be equal to the relative weight of each measurement in

the regression vector. The new integration times, \mathbf{b}^\dagger , are determined by the regression vector,

$$\mathbf{b}^\dagger = \mathbf{b}(J/\|\mathbf{b}\|_1). \quad (10.22)$$

where $\|\mathbf{b}\|_1$ is the one-norm, or sum of the absolute values. Equation (10.22) sets the time dedicated to integration at each channel to be from zero to J time units. Some channels, the most influential ones for regression, will be allocated more than the 1.0 time units given by normal digitization, and the least important channels can, in the limit, be eliminated if the regression coefficient is zero. The standard deviation of estimation for OR is determined to be

$$\sigma_{\hat{c}} = \left(\sum_{j=1}^J \sigma_j^2 \right)^{1/2} \|\mathbf{b}\|_1/J = J^{1/2} \sigma \|\mathbf{b}\|_1/J = \sigma \|\mathbf{b}\|_1/J^{1/2} \quad (10.23)$$

Comparison of Equation (10.23) to Equation (10.21) shows the OR will yield more precise estimations whenever

$$2^{1/2} \sigma \|\mathbf{b}\|_1/J < \sigma \|\mathbf{b}\|_2 \quad (10.24a)$$

or, equivalently,

$$2^{1/2}/J^{1/2} < J^{1/2} \|\mathbf{b}\|_2/\|\mathbf{b}\|_1 \quad (10.24b)$$

Since, by mathematical proof $J^{1/2} \|\mathbf{b}\|_2/\|\mathbf{b}\|_1 \geq 1$ and J is 2 or greater by definition for multivariate analysis, OR is expected to yield lower standard deviations of estimation than DR.

10.6.2 APPLICATION

The predictive performance of OR and DR are compared on a collection of tertiary mixtures of dense nonaqueous liquids. NIR transmittance spectra are collected with a low-resolution, fiber-optic NIR spectrometer. A quartz tungsten halogen lamp was used as the light source (Oriel). This light was focused on an optical fiber with a 0.38 numerical aperture. This fiber was connected to an Ocean Optics NIR/visible transmission cell. The transmitted light was collected by a second optical fiber. An acousto-optic tunable filter (AOTF, Brimrose) was employed for wavelength discrimination and a thermoelectrically cooled InGaAs detector (Oriel) was employed for signal transduction. The AOTF and InGaAs detector were controlled by Labview software. A three-level mixture design (Figure 10.10) was employed to construct tertiary mixtures of dichloromethane, toluene, and cyclohexane. Spectra were collected from 1105 to 1200 nm with 5 nm resolution. Each wavelength was integrated for 6.25 sec at 80 Hz for the calibration spectra. Spectra of the pure constituents are presented in Figure 10.11.

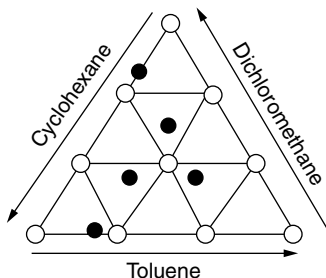


FIGURE 10.10 Experimental design for comparison of OR and DR. Light gray circles designate standards, black circles designate unknowns.

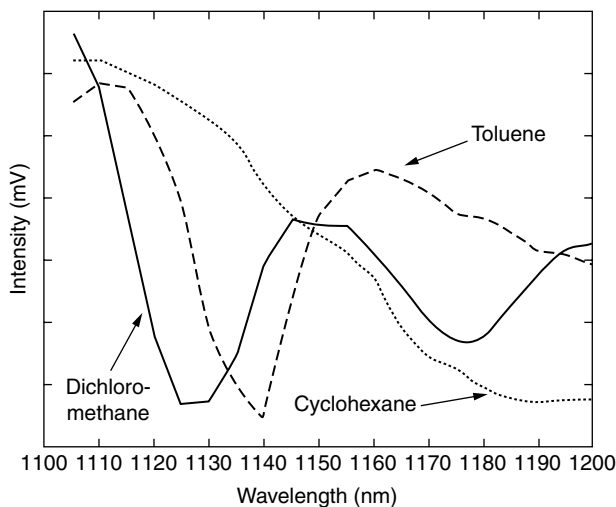


FIGURE 10.11 Recorded NIR transmittance spectra of dichloromethane (solid), toluene (dashed), and cyclohexane (dotted).

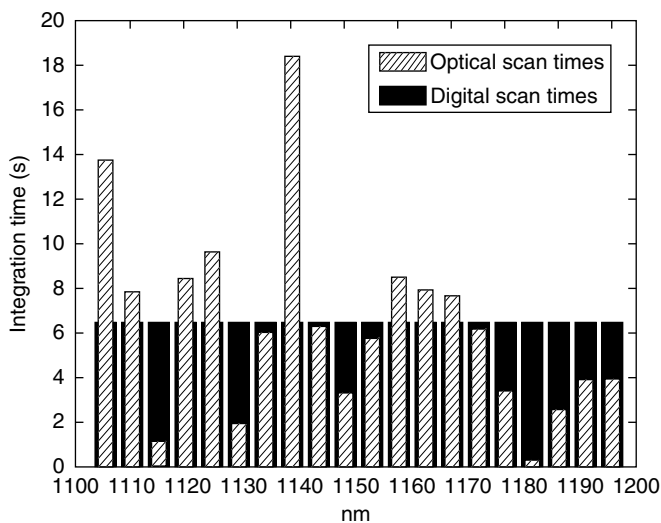


FIGURE 10.12 Integration times for DR (solid) and OR (striped) applied to quantitation of dichloromethane in a mixtures with toluene and cyclohexane.

A comparison of the relative integration times at each wavelength for OR and DR is shown in Figure 10.12 for dichloromethane. Plots for toluene and cyclohexane appear similar. The integration time at each wavelength for OR is proportional to the absolute value of the regression vector at the specific wavelength with the total integration time scaled to the total integration time allocated for DR. Consequently, with the OR scaling, the most analytically important wavelengths will have the best signal-to-noise.

As predicted by theory, OR returns more precise estimates of analyte concentration than DR. Figure 10.13 presents the plot of precision of analysis for the five unknowns described in Figure 10.10. Each standard deviation is calculated from ten replicated analyses. The accuracies of the two alternatives are equal because both methods employ the same regression vector. The relative precision

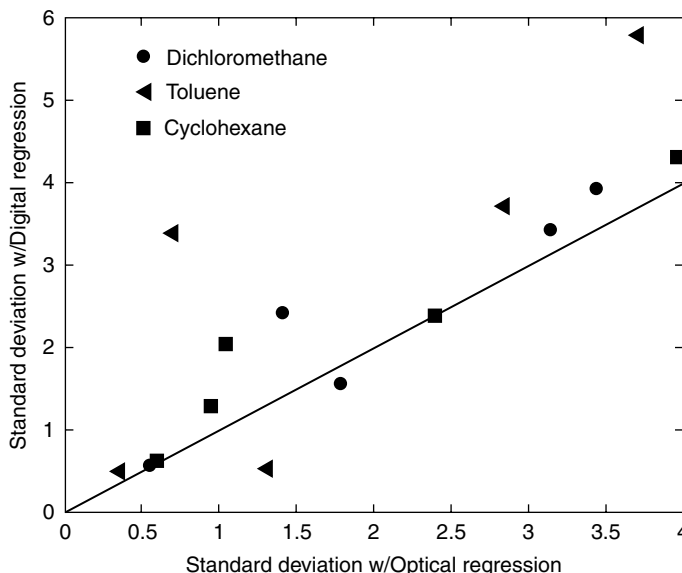


FIGURE 10.13 Comparison of precision of estimation for OR and DR applied to NIR transmittance of dichloromethane, toluene, and cyclohexane.

of estimation varies among the 15 samples based on location in the experimental design. Samples farthest from the center of the experimental design are expected to have larger standard deviations of estimation. Two of the fifteen concentration estimates are less precise for OR than for DR. This is primarily due to random chance with the small number of replicate samples analyzed. However, taken as a group, analysis by the paired difference test shows that OR is more precise than DR at greater than the 99% confidence level.

ACKNOWLEDGMENTS

The authors thank Dr. Mike Blackburn of Cargill for the NIR reflectance data of corn flour. The authors wish also to thank the National Science Foundation and Center for Process Analytical Chemistry for supporting the OR investigations.

REFERENCES

1. P. Geladi, D. MacDougall, and H. Martens, "Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat," *Appl. Spectros.*, **39**: 491–500 (1985).
2. T. Isaksson and T. Naes, "The Effect of Multiplicative Scatter Correction and Linearity Improvement on NIR Spectroscopy," *Appl. Spectros.*, **42**: 1273–1284 (1988).
3. T. Isaksson and B. Kowalski, "Piece-Wise MSC Applied to Near-Infrared Diffuse Transmittance Data from Meat Products," *Appl. Spectros.*, **47**: 702–709 (1993).
4. S. Wold, H. Antti, F. Lindgren, and J. Ohman, "Orthogonal Signal Correction of Near Infrared Spectra," *Chemom. Intell. Lab. Syst.*, **44**: 175–185 (1998).
5. S. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, and S. Wold, "An Evaluation of Orthogonal Signal Correction Applied to Calibration Transfer of Near Infrared Spectra," *Chemom. Intell. Lab. Syst.*, **44**: 229–244 (1998).
6. T. Fearn, "On Orthogonal Signal Correction," *Chemom. Intell. Lab. Syst.*, **50**: 47–52 (2000).
7. Y. Wang, M. J. Lysaght, and B. R. Kowalski, "Improvement of Multivariate Calibration through Instrument Standardization," *Anal. Chem.*, **64**: 562–565 (1992).

8. C. S. Chen, C. W. Brown, and S. C. Lo, "Calibration Transfer from Sample Cell to Fiber-Optic Probe," *Appl. Spectros.*, **51**: 744–748 (1997).
9. J. Lin, "Near-IR Calibration Transfer between Different Temperatures," *Appl. Spectros.*, **52**: 1591–1605 (1998).
10. P. J. Gemperline, J. H. Cho, P. K. Aldridge, and S. S. Sekulic, "Appearance of Discontinuities in Spectra Transformed by the Piecewise Direct Instrument Standardization Procedure," *Anal. Chem.*, **68**: 2913–2915 (1996).
11. T. B. Blank, S. T. Sum, S. D. Brown, and S. L. Monfre, "Transfer of Near-Infrared Multivariate Calibrations without Standards," *Anal. Chem.*, **68**: 2987–2995 (1996).
12. S. T. Sum and S. D. Brown "Standardization of Fiber Optic Probes for Near Infrared Multivariate Calibrations," *Appl. Spectros.*, **52**: 869–877 (1998).
13. K. S. Booksh and B. R. Kowalski, "Theory of Analytical Chemistry," *Anal. Chem.*, **66**: 782A–791A (1994).
14. I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.
15. J. M. Sutter, J. M. Kalavias, and P. M. Lang, "Which Principal Components to Utilize for Principal Components Regression," *J. Chemom.*, **6**: 217–225 (1992).
16. A. Lorber, L. E. Wangen, and B. R. Kowalski, "A Theoretical Foundation for the PLS Algorithm," *J. Chemom.*, **1**: 19–31 (1987).
17. P. Geladi and B. R. Kowalski, "PLS Regression: A Tutorial," *Anal. Chim. Acta*, **185**: 1–17 (1986).
18. R. Marbach and H. M. Heise, "On the Efficiency of Algorithms for Multivariate Linear Calibration used in Analytical Spectroscopy," *TRAC*, **11**: 270–275 (1992).
19. M. Stone and R. J. Brooks, "Continuum Regression: Cross-validated Sequentially-constructed Prediction Embracing Ordinary Least Squares, Partial Least Squares, and Principal Component Regression," *J. R. Stat. Soc. B.*, **52**: 337–369 (1990).
20. B. M. Wise and B. R. Kowalski, Process Chemometrics, in *Process Analytical Chemistry* (F. McLennan and B. R. Kowalski, eds.), Blackie Academic, London, 1995.
21. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, and L. Kaufman, *Chemometrics: A Textbook*, Elsevier, Amsterdam, 1988.
22. D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. DeJong, P. J. Lewi, and J. Smeyers, *Handbook of Chemometrics and Qualimetrics*, Elsevier, Amsterdam, 1997.
23. S. de Jong, "SIMPLS: An Alternative Approach to Partial Least Squares Regression," *Chemom. Intell. Lab. Syst.*, **18**: 251–263 (1993).
24. R. Manne, "Analysis of Two PLS Algorithms for Multivariate Calibration," *Chemom. Intell. Lab. Syst.*, **2**: 187–197 (1987).
25. T. Naes and T. Isaksson, "Locally Weighted Regression of Diffuse Near Infrared Transmittance Spectroscopy," *Appl. Spectros.*, **46**: 34–43 (1992).
26. T. Naes, T. Isaksson, and B. R. Kowalski, "Locally Weighted Regression and Scatter Correction for Near-Infrared Reflectance Data," *Anal. Chem.*, **62**: 664–673 (1990).
27. Z. Wang, T. Isaksson, and B. R. Kowalski, "New Approach for Distance Measurement in Locally Weighted Regression," *Anal. Chem.*, **66**: 249–260 (1994).
28. E. Malinowski, *Factor Analysis in Chemistry*, 2nd edn, John Wiley & Sons, New York, 1991.
29. A. M. C. Prakash, C. M. Stellman, and K. S. Booksh, "Optical Regression: A Method for Improving Quantitative Precision of Multivariate Prediction with Single Channel Spectrophotometers," *Chemom. Intell. Lab. Syst.*, **46**: 265–274 (1999).

11 Transfer of Multivariate Calibration Models Based on Near-Infrared Spectroscopy

Eric Bouveresse and Bruce Campbell

CONTENTS

11.1	Introduction	232
11.1.1	Near-Infrared Spectrophotometry in Analytical Chemistry	232
11.1.2	Development of NIR Calibration Models	232
11.1.3	Situations Yielding Invalid Predictions	233
11.1.4	How to Deal with Such Situations?.....	233
11.2	Standardization Methods	234
11.2.1	Principles of Standardization Methods	234
11.2.2	Selection of Standardization Samples	234
11.2.2.1	Stability and Representativity of the Standardization Samples	234
11.2.2.2	Approaches for Selecting Standardization Samples	234
11.2.2.3	Selection of the Most Suitable Approach	235
11.2.2.4	Number of Standardization Samples	236
11.2.3	Computation of the Standardization Parameters	237
11.2.3.1	The Different Approaches to Compute Standardization Parameters ...	237
11.2.3.2	Standardization Methods Based on Correcting Predicted Values.....	237
11.2.3.3	Standardization Methods Based on Transferring NIR Spectra.....	237
11.2.3.4	Standardization Methods Based on Transferring Calibration Models ..	239
11.2.3.5	Validation of the Standardization Parameters.....	239
11.2.3.6	Forward Transfer.....	240
11.3	Development of Robust Calibration Models	240
11.3.1	Models Robust against Expected Sources of Variations.....	240
11.3.2	Models Robust against Simulated Perturbations	240
11.3.3	Robust Models Based on Pretreated Spectra	241
11.4	Instrument Recalibration	241
11.5	Conclusion	242
	References	242

11.1 INTRODUCTION

11.1.1 NEAR-INFRARED SPECTROPHOTOMETRY IN ANALYTICAL CHEMISTRY

In many different fields of analytical chemistry, there is a constant need for more and more performant analysis methods. To be performant, such methods should allow an easy, fast and cheap analysis of a wide variety of products. Therefore, near-infrared (NIR) spectrophotometry has become one of the most powerful techniques in analytical chemistry [1–5], because of the following advantages:

1. NIR spectrophotometry enables the analysis of a wide variety of samples, including for instance strongly absorbing samples or opaque solid materials.
2. NIR wavelengths allow the use of long fiber optics, which is very convenient for online process analysis.
3. NIR spectrophotometry requires little or no sample preparation and therefore enables easy and fast data collection. It also decreases or eliminates errors due to sample preparation.
4. NIR spectrophotometry is much faster than the majority of the classical analytical techniques. Indeed, the analysis time ranges from a few seconds to several minutes [2].
5. NIR spectrophotometry is a nondestructive method. This represents a considerable advantage for applications where the analyzed samples must not be altered (e.g., biological and medical applications [6], online process monitoring, etc.).
6. NIR spectrophotometry enables the determination of several properties from a single measurement.

Because of these considerable advantages, NIR spectrophotometry has become more and more popular in many different industrial fields, such as the agroalimentary industry [5], the pharmaceutical industry [7–10], the chemical industry [11,12], the petroleum industry [13,14], the cosmetic industry [15], the textile industry [16], etc. Overviews of NIR applications are available in the literature [1–3].

11.1.2 DEVELOPMENT OF NIR CALIBRATION MODELS

The development of NIR calibration models consists of four steps:

1. The first one is the analysis of the calibration samples by the reference method to obtain the corresponding reference value of the studied parameter (e.g., concentration of a particular compound) for each calibration sample.
2. The second one is to obtain the NIR spectrum of each calibration sample on a NIR instrument.
3. The third one is the determination of the mathematical relationship between the NIR spectra and the corresponding reference values of the studied parameter (the y-value).
4. The fourth one is the validation of the calibration with independent samples. These validation samples are analyzed by both NIR and reference techniques to check that both are in acceptable statistical agreement.

The validated calibration model can then be used to predict values of the studied parameter for new samples. The process is to scan the new samples with the NIR instrument and use the calibration to predict the studied parameter. It should be observed that correct predictions could only be obtained if the NIR spectra of the new samples are within the range of variables (concentration, sample temperature, sample age, etc.) covered by the calibration samples. Because extrapolation of the calibration model outside these ranges almost always leads to erroneous predictions, it is usually more convenient to develop calibration models on very large databases, in order to cover the expected number of sources of variations. In practice, calibration databases can sometimes involve several hundred, if not thousands, of calibration samples collected over very long periods of time. Therefore, the development of such calibration models requires considerable effort, high costs, and long delays.

11.1.3 SITUATIONS YIELDING INVALID PREDICTIONS

Calibration models can yield invalid predictions when changes occur between the measurements of the calibration samples (referred to as the *calibration step* in this chapter) and the measurements of new samples for prediction (referred to as the *prediction step* in this chapter). Three main sources of changes can be distinguished:

The first source is that the instrumental response of a spectrophotometer can materially change between the calibration and prediction steps. The instrumental response of a NIR instrument can be subject to variations due to instrument aging (continuous drift in the instrumental response) or to an important repair (sudden shift in the instrumental response). If the instrumental response of the spectrophotometer changed between the calibration and the prediction step, incorrect predictions will be obtained.

The second source is in the situation where the calibration is done on one spectrophotometer, the primary or master one, and is transferred to a second (the slave, secondary, server, host) spectrophotometer. Without a proper transfer, erroneous predictions will be obtained. The reason for this is each NIR spectrophotometer has its own instrumental response, and this response is different from the one of another instrument, even if both instruments are identical (same characteristics, same manufacturer, same model).

The last source arises from variations in the measurement conditions between both calibration and prediction steps (changes in temperature, changes of measurement cell, etc.). For instance, physical parameters such as temperature can have a strong influence on the bond vibrations, and therefore on the shape of the resulting NIR spectra [17]. If the temperature or temperature range of the predicted samples is not identical to that of the measured samples in the calibration step, predictions obtained with the spectra collected in the prediction step will be erroneous.

Since the predictive ability of a calibration model can be seriously damaged by the previously described sources of changes, it means that a calibration model developed in well-defined measurement conditions on an NIR instrument will provide reliable predictions only if new samples are measured using the same measurement conditions and on the same instrument, the instrumental response of which has to be nearly constant. In many practical situations, it is very difficult to fulfill all these requirements. However, since the development of a calibration model is a long, sometimes tedious, and delicate procedure, NIR users would like to avoid developing separate calibration models on each instrument. This implies the measurement of numerous calibration samples on each instrument. If the instrumental response of the NIR spectrophotometer fluctuates unacceptably over time, the spectrophotometer must be periodically recalibrated. In this case, at a minimum, calibration samples would have to be rescanned. If the calibration samples are chemically unstable samples, which simply cannot be stored over time, this recalibration must be performed with new calibration samples. Additional work, time, and efforts are necessary to analyze these new calibration samples by both reference and NIR methods; the work performed for the analysis of older calibration samples being no longer useful. Finally, if the measurement conditions of the new samples have changed, the instrument must be recalibrated with calibration samples remeasured in the new measurement conditions, which again implies periodic remeasurement of numerous calibration samples.

11.1.4 HOW TO DEAL WITH SUCH SITUATIONS?

To avoid situations yielding invalid predictions resulting from the problems outlined previously, different approaches have been proposed:

Standardization methods: In order to correct calibrations that no longer yield valid predictions, standardization procedures can be used [18–21]. These procedures are based on three main steps.

The first step consists of scanning some well-chosen samples (referred to as *standardization samples*) in both calibration and prediction steps, in order to estimate the differences between both calibration and prediction steps. The second step consists of computing standardization parameters, in order to correct the estimated differences. The choice of the standardization samples and the method used to compute the standardization parameters have to be carefully studied and chosen in order to obtain an optimal standardization. Finally, the standardization coefficients must be validated with independent samples measured in both calibration and prediction steps.

Development of robust calibration models: This approach consists of building calibration models that are not sensitive to the spectral differences due to the previously mentioned changes. Even if spectral differences occur between both calibration and prediction steps, the use of such robust calibration models yields standardized predicted y-values.

Instrument recalibration: This approach consists of selecting a subset of calibration samples representative of the original calibration data set. When the current calibration model does not yield good predictions, this subset can be remeasured on the instrument, and a new calibration model is determined with the new spectra of the subset samples.

11.2 STANDARDIZATION METHODS

11.2.1 PRINCIPLES OF STANDARDIZATION METHODS

The three important steps of standardization methods are (a) estimation of the spectral differences between calibration and prediction steps, (b) calculation of standardization parameters able to correct the estimated differences, and (c) validation of the standardization parameters.

11.2.2 SELECTION OF STANDARDIZATION SAMPLES

To estimate the differences between calibration and prediction steps, standardization samples are measured in both steps. The choice of the standardization samples is therefore very important, and particular attention must be paid to this delicate problem. In this section, the different types of standardization samples are presented and the main criteria for selecting standardization samples are discussed, namely the stability and representativity of the standardization samples.

11.2.2.1 Stability and Representativity of the Standardization Samples

Stability: The physical and chemical stability of the standardization samples is the crucial point for successful standardization. Indeed, if changes occur in the standardization samples between both calibration and prediction steps, a part of the spectral differences will be due to these changes, and it will not be possible to differentiate these differences from the ones between calibration and prediction steps. The resulting standardization parameters will therefore correct the instrumental differences but will also incorporate the differences due to the instability of the standardization samples. Such standardization parameters will yield erroneous results when used for predictions in routine practice.

Representativity: Standardization samples should be as representative as possible of the samples to which the standardization parameters will be applied. If the standardization samples are not representative, the differences seen in the use of them will usually not be germane to the intended use. Thus, erroneous correction factors may be derived; factors that will cause future predictions to be in error. The further the standardization samples are from representative, the worse the situation.

11.2.2.2 Approaches for Selecting Standardization Samples

Two main approaches can be distinguished; selection of a representative subset of standardization samples among the calibration samples and the use of independent standardization samples.

Selection of a representative subset consists of selecting a subset of representative samples among the numerous calibration samples collected in the calibration step and measuring the samples from this standardization subset in the prediction step. The main advantage of the standardization subset is the high representativity of the standardization samples, provided a good subset selection method is used. Different subset selection methods have been proposed in the literature [22,23]. A comparative study [24] showed that the smallest and most representative subset is obtained by the Kennard and Stone algorithm [23]. One important consideration is to ensure samples in the subset cover the entire range of values.

Selection of independent standardization samples consists of measuring a set of independent samples on both instruments. These independent samples are not used for model development. Either samples close to the physical and chemical structure of the analyzed samples or very different samples such as generic standards can be used. The main advantage of this approach is the user does not have to develop special storage conditions that probably still would not ensure having stable standardization samples.

11.2.2.3 Selection of the Most Suitable Approach

The selection of the most suitable approach consists of making the best compromise between representativity and stability for a given situation if both cannot be perfectly met. The use of a representative subset of calibration samples can yield representative standardization samples, but the stability of the samples may be a problem (e.g., for food products). By using independent standardization samples, it is possible to choose samples stable enough between both calibration and prediction steps, but these samples are less representative. This compromise can be summarized by the plot shown in Figure 11.1. The next part of this chapter indicates which approach should be chosen in a number of given situations.

When the calibration samples are very stable or can be kept stable over time (e.g., by protecting samples from the light, temperature effects, loss of volatile components, and by storing them under inert atmosphere), the subset selection approach should be used, because it enables one to obtain stable and representative standardization samples. This situation is the ideal one, because there is no real compromise to be made between stability and representativity.

During the measurement of the calibration samples, a set of independent standardization samples has to be simultaneously measured to check the stability of the instrumental response. For the standardization of a single instrument over time, the stability is the most critical point since the same

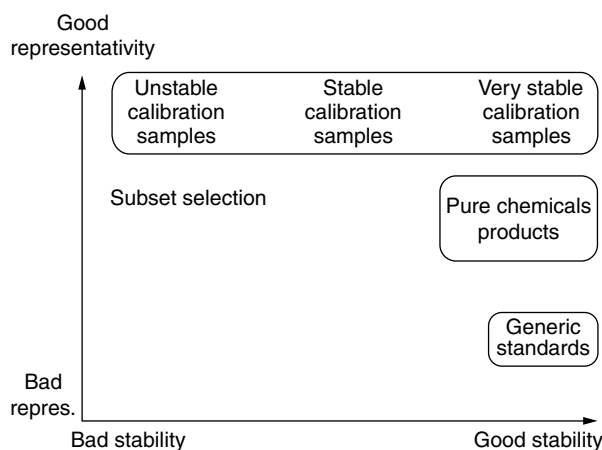


FIGURE 11.1 Stability and representativity relationship of the standardization samples.

samples should be used over very long periods of time. In such a case, only perfectly stable samples can be used as standardization samples. If the analyzed samples are or can be kept perfectly stable, a representative standardization subset can be selected among the calibration samples. However, if they cannot be kept stable over time, it is necessary to select independent stable standardization samples. These should be as representative of the samples to be analyzed as possible. The most stable standardization samples are generic standards such as halons [25–27], polystyrene [28], etc. However, these samples are usually so different from the analyzed samples, that the results obtained are usually not satisfactory. Therefore, the use of these standards should be limited to the monitoring of particular instrumental parameters over time [29,30]. A more suitable way of obtaining a stable and representative standardization set is to use a set of mixtures containing well-chosen pure chemicals similar to the analyzed samples [31]. However, even this may lead to error, as slight, unknown amounts of impurities can differ between bottles of the same production lot. One such example is the presence of particulates in containers used to store the chemicals.

For the standardization of different instruments, it should be noted that the stability of the standardization samples is not as critical as for the standardization of a single instrument over time. When two NIR instruments must be standardized, the standardization samples need only be stable between the moment at which they are measured on the master instrument and the moment at which they are measured on the slave instrument. This period of time does not exceed a few minutes, if both instruments are located at the same place, and is not longer than a few days, if both instruments are far from each other. Therefore, less stable but more representative standardization samples can be used. For instance, Shenk and Westerhaus provide a set of 30 standardization cups containing different agricultural products that are representative of most of the agricultural samples currently analyzed by NIR spectrophotometry [32,33]. These agricultural samples are probably not perfectly stable over several years, but because they can be considered as very stable over the space of a few days, they can therefore be used as standardization samples.

11.2.2.4 Number of Standardization Samples

A sufficiently high number of samples must be included in the standardization set in order to contain enough information concerning the instrumental differences between both calibration and prediction steps. Too low of a number of standardization samples will yield an incomplete standardization. Too high of a number of standardization samples represents a loss of time. There are no absolute rules to decide how many standardization samples should be used, but this number is strongly conditioned by two factors, namely the standardization method used and the complexity of the spectral differences. The choice of how many standardization samples often depends on the standardization method and complexity of the spectra.

The number of standardization samples depends on the standardization approach used. For instance, the method proposed by Shenk and Westerhaus [32,34] is based on a wavelength index correction followed by a spectral intensity correction. To obtain a performant wavelength index correction step, the ISI standardization procedure requires 15 to 30 standardization samples. When the piecewise direct standardization (PDS) is applied [22,28,35–37], it is recommended to select a subset of calibration samples. Due to the local correction of the PDS algorithm, the resulting number of standardization samples is often smaller than 10, and satisfactory results were obtained in most of the cases reported [24].

The number of standardization samples also depends on the complexity of the differences between both calibration and prediction steps. For instance, Shenk and Westerhaus suggested using 30 standardization samples in order to standardize instruments that require wavelength corrections [33]. However, for NIR instruments that do not need such wavelength corrections, Shenk and Westerhaus introduced the “single sample standardization,” which consists of measuring a single standardization sample in both calibration and prediction steps and correcting the spectral intensities by the offset resulting from the spectral differences of the single standardization sample [33].

11.2.3 COMPUTATION OF THE STANDARDIZATION PARAMETERS

11.2.3.1 The Different Approaches to Compute Standardization Parameters

Different approaches have been proposed to compute standardization parameters used to correct the differences between both calibration and prediction steps. Indeed, standardization parameter can be determined in order to correct the predicted y -values, the NIR spectra, or the calibration coefficients.

11.2.3.2 Standardization Methods Based on Correcting Predicted Values

This standardization approach (usually referred to as “the slope/bias correction”) consists of computing predicted y -values for the standardization samples with the calibration model. These transfers are most often done between instruments using the same dispersion device, in otherwords, Fourier transform to Fourier transform, or grating to grating. The procedure is as follows. Predicted y -values are computed with the standardization spectra collected in both calibration and predicted steps. The predicted y -values obtained with spectra collected in the calibration step are then plotted against those obtained with spectra collected in the prediction step, and a univariate bias or slope/bias correction is applied to these points by ordinary least squares (OLS). For new spectra collected in the prediction step, the calibration model computes y -values and the obtained predictions are corrected by the univariate linear model, yielding standardized predictions.

The main advantage of this method is that it requires only a simple univariate correction. This approach is simple, fast, and easy to understand. Therefore, many applications using the slope/bias correction as standardization method can be found in the literature [38–40].

If several properties must be predicted from a single NIR spectrum, this method should be applied for each property, in order to correct each calibration model independently. Moreover, the slope/bias correction requires the computation of y -values for the standardization samples, which makes the use of generic standards impossible. Finally, this simple approach can only be used for instrument standardization when the differences between calibration and prediction steps are rather simple. When more complex instrumental differences occur, the predictions cannot be corrected by a simple univariate correction. To decide whether the slope/bias correction method can be successfully applied, a procedure based on a statistical F-test was proposed [41].

11.2.3.3 Standardization Methods Based on Transferring NIR Spectra

This approach consists of transforming the spectra of the new samples collected in the prediction step to be similar to the spectra used for the calibration step. The calibration model can therefore be applied to these transformed spectra in order to obtain correct predictions of the studied properties.

A number of standardization methods based on transferring NIR spectra were proposed in the literature, namely a patented algorithm proposed by Shenk and Westerhaus [32–34], the two-block partial least-squares (PLS) algorithm [42,43], the direct and PDS algorithm [22], standardization algorithms using the wavelet domain [44], and standardization algorithms using artificial neural networks [45,46], etc. The aim of this chapter is not to give a detailed description of all existing standardization algorithms [18–21], but to present the specificity of the algorithms based on transferring spectra, their advantages, and their limitations. The aim of this chapter is also to present a way of selecting the most suitable algorithm for transferring NIR spectra. The next questions can be very helpful in deciding which algorithm is the most suitable.

Type of the spectrophotometers involved? When considering a transfer, the type of spectrophotometer should be considered. Transfers between the same model of instrument from the same

manufacturer are usually the easiest. When transfers between different models are attempted, the difficulty increases. Hardest of all are the ones between instruments that have different approaches to the dispersion of the radiation, such as between Fourier transform-based and wavelength dispersion.

Do NIR spectra have the same data resolution? The first essential point of a standardization problem concerns the type of spectrophotometers involved. The majority of the algorithms existing in the field of instrument standardization are able to deal with instruments providing exactly the same numbers of data points at the same wavelengths. However, it can happen that instruments of different types need to be standardized. This is, for instance, the case when a calibration model developed on an expensive high-resolution monochromator (e.g., in a central laboratory) has to be transferred to a cheap filter instrument (e.g., in an industrial process). Among the standardization algorithms previously mentioned, only two methods can be applied, namely two-block PLS [43,44] and direct standardization [22]. These two methods first compress the information in a few principal components (for direct standardization) or in a few latent variables (for two-block PLS), and then compute a transfer matrix relating both computed data matrices. The major advantage of these two methods is that calibration models developed on the monochromator instrument can directly be transferred to the filter instrument (see Section 11.2.3.4). However, parameters such as the number of principal components or latent variables to be used have to be optimized.

Another problem of data resolution is that a number of standardization algorithms can only be applied to spectra made of a regularly sampled data sequence (monochromators, Fourier-transform instruments, etc.), but they cannot be used with spectra made only of a few discrete wavelengths (filter instruments). For instance, the PDS algorithm based on a moving window of neighboring wavelengths [22] or the standardization in the wavelet domain requiring a regularly sampled data sequence [42] cannot be used, if the slave instruments used are filter based.

Are the standardization samples representative of the calibration samples? The second important point is the nature of the standardization samples used. For some standardization algorithms, successful standardization results can only be obtained if the representativity of the standardization samples is satisfactory. This is the case for multivariate algorithms such as the two-block PLS [43,44], the direct and PDS [22] or the algorithms based on neural networks [45,46]. Therefore, these algorithms can only be applied when the experimental calibration domain is well covered by the standardization samples [24].

For these algorithms based on local multivariate corrections, a poor representativity of the standardization at a few wavelengths can create artifacts in the transferred spectra. It should be noted that the local but univariate spectral intensity correction of the Shenk–Westerhaus algorithm is less sensitive to artifacts. When less representative standardization samples are used, ones that cover a range much larger than the one covered by the calibration samples, nonlinearities in the spectral intensities can appear. To deal with these nonlinearities, the Shenk–Westerhaus algorithm has been modified using locally weighted regression [31]. They reported acceptable results with pure chemicals as standardization samples.

How many standardization samples are available for standardization? An important point to consider is the number of standardization samples necessary to estimate the transfer parameters. Some methods based either on univariate local corrections (Shenk–Westerhaus algorithm) or on multivariate local correction (PDS, standardization in the wavelet domain, etc.) can be performed with a reduced number of standardization samples. However, multivariate methods based on the whole spectral range (direct standardization, two-block PLS) usually require more standardization samples to obtain a good and stable estimation of the standardization parameters [21].

How complex are the spectral differences? Another point to be checked is the complexity of the instrumental differences. The instrumental differences existing between NIR spectrophotometers can be either very simple (e.g., global systematic bias in spectral intensities, etc.) or more complex (e.g., local band broadening, local peak shift, etc.).

If two instruments from the same type and from the same manufacturer are used, one can expect that the spectral differences will be easier to correct, and one will therefore select a simple method able

to deal with such simple differences. In such a case, one would recommend the Shenk–Westerhaus algorithm (simple, univariate, less sensitive to artifact, etc). However, if different instruments of different quality and from different manufacturers are used, or if there are temperature variations between the calibration and prediction steps, one can expect more complex spectral differences, and one will apply a more powerful standardization algorithm able to deal with such changes. For instance, when instrumental differences between calibration and prediction steps become more complex (e.g., peak broadening), the univariate corrections used in the Shenk–Westerhaus algorithm are not able to deal with such differences anymore, and incorrect results can be obtained. In such a case, the use of local and multivariate standardization methods should be recommended, such as the PDS method. Another weakness of the Shenk–Westerhaus algorithm is its linear character. If the range covered by the standardization samples is small, spectral intensities can be successfully corrected by univariate linear regressions. However, when the standardization samples cover a larger spectral intensity range, nonlinear effects due to samples with large spectral intensity values can occur.

In all cases, it is highly recommended to perform a visual inspection of the spectral differences before deciding on the standardization method to be used. If spectral changes differ substantially from one wavelength range to another one, local standardization methods such as the PDS or the standardization in the wavelet domain is recommended. It should be noted that the standardization in the wavelet domain should be preferred when the signal-to-noise level of the slave instrument is much lower than the one of the master. This is because of the ability of wavelets to decrease or even eliminate noise from spectral data. Finally, if spectral changes show significant nonlinearities, the standardization using neural networks can be recommended.

11.2.3.4 Standardization Methods Based on Transferring Calibration Models

This standardization approach consists of transferring the calibration model from the calibration step to the prediction step. This transferred model can be applied to new spectra collected in the prediction step in order to compute reliable predictions. An important remark is that the standardization parameters used to transfer calibration models are exactly the same as the ones used to transfer NIR spectra. Some standardization methods based on transferring spectra yield a set of transfer parameters. For instance, the two-block PLS algorithm yields a transfer matrix, and each new spectrum collected in the prediction step is transferred by simply multiplying it by the transfer matrix. For these standardization methods, the calibration model can be transferred from the calibration step to the prediction step using the same transfer matrix. It should be pointed out that all standardization methods yielding a transfer matrix (direct standardization, PDS, etc.) could be used in order to transfer the model from the calibration to the prediction step. For instrument standardization, the transfer of a calibration model from the master instrument to the slave instruments enables each slave instrument to compute its own predictions without systematically transferring the data back to the master instrument.

11.2.3.5 Validation of the Standardization Parameters

To validate the standardization parameters, independent samples should be measured in the prediction step. Depending on the retained approach, either the resulting spectra should be transferred, or the transferred calibration model should be applied to the spectra. Standardized predictions are obtained and these predictions must be compared either with the reference values (this implies that the standardization samples have been analyzed by the reference method), or with the predictions obtained in the calibration step (this implies that the standardization samples have been measured in the calibration step). This procedure verifies the standardized predictions are in good statistical agreement with either the reference method or the original calibration model.

11.2.3.6 Forward Transfer

The forward transfer involves developing a transform process for altering spectra from the calibration step to the prediction step. The transformed spectra can then be used to develop a calibration in the prediction step. This process is much faster than a calibration done using new calibration standards. The transform process can be done with only a few standards scanned in both steps, and because this is done over a short time interval, the stability of standards is usually not a problem.

The forward transfer is particularly interesting in the following situations:

1. A more performant instrument has been bought to replace an old one. The whole calibration database should be transferred from the old to the new instrument, and the calibration model should be developed with the transformed spectra.
2. The instrumental response of a single NIR instrument has changed over time. Instead of correcting the collected spectra of new samples, it is also possible to transfer the whole calibration database to match the new conditions.
3. Process conditions such as temperature have been modified and the new samples are not within the calibration domain anymore. Instead of correcting the collected spectra of new samples for temperature, it is also possible to transfer the whole calibration database to account for the spectral changes due to temperature variations.

11.3 DEVELOPMENT OF ROBUST CALIBRATION MODELS

This approach consists of building calibration models that are not sensitive to the spectral differences due to the previously mentioned changes. The main advantage of robust calibration models is that they can be transferred without performing any correction. Therefore, no standardization samples need to be measured in both calibration and prediction steps.

11.3.1 MODELS ROBUST AGAINST EXPECTED SOURCES OF VARIATIONS

An interesting approach yielding robust calibration models was proposed by De Noord [18]. It consists of incorporating in the calibration design all sources of variation that can be expected. If one wishes to develop a calibration model robust against temperature variations and differences between instruments, one will measure calibration samples at different temperatures and on different instruments. Extracting the information relevant for modeling from such a calibration set yields a calibration model that is orthogonal to the included sources of variations. It can therefore be applied to new samples, whatever the sample temperature and the instrument used.

This approach is very powerful, and the resulting model is not sensitive to the sources of variations included in the design. However, the main limitation of this approach is that it is very difficult to foresee all sources of variation beforehand [18,47]. Erroneous results will be obtained if unexpected sources of variation appear. In such a case, this new source of variation has to be included in the calibration database and the calibration model must be updated. Such an approach can be recommended only when a calibration model is developed on a limited number of instruments, otherwise the amount of data to be collected would be too large.

11.3.2 MODELS ROBUST AGAINST SIMULATED PERTURBATIONS

This approach consists of selecting wavelengths that are robust against the spectral differences between the calibration and prediction steps and using these selected wavelengths to compute calibration coefficients. This approach was proposed by Mark and Workman [48] to build a calibration model

robust against wavelength shifts between two NIR instruments. To determine which wavelength is robust, small variations of the wavelengths are introduced and the resulting changes in the regression coefficients are studied. The wavelengths for which the wavelength variations have no effect on the regression coefficients are selected and used to build robust MLR models. Such an approach was also used by Despagne et al. [49], where the PLS calibration model the most robust against simulated perturbations (wavelength shift, linear differences in spectral intensities, and stray light, etc.) was retained.

This method is very efficient in building models robust against perturbations that are known in advance and can be simulated. Moreover, the simulation of the perturbations enables avoiding the measurement of large data sets under many different conditions. However, the main limitation of this approach is that the model can be made robust only against variations that can be simulated. Erroneous results will be obtained if sources of variation appear that cannot be simulated. In such a case, this new source of variation has again to be included in the calibration database, and the model must be updated. Such a calibration model should be recommended when most of the expected variations can be simulated. This is the case when one wishes to develop a calibration model robust against fluctuations of the instrumental response over time. This enables the reduction of the number of situations that need either standardization over time or model updating.

11.3.3 ROBUST MODELS BASED ON PRETREATED SPECTRA

This approach consists of eliminating all sources of variation not related to the y -values through data preprocessing. The elimination of information useless for the modeling leads to more robust calibration models. This approach was applied to build PLS models for the NIR of heavy oil products [47], and it was demonstrated that using the calibration model based on spectra after a second derivative mathematical treatment followed by a multiplicative signal correction was less sensitive to wavelength drifts than the calibration model based on raw spectra. Other preprocessing techniques, such as the orthogonal signal correction (OSC) [50,51], can be used to eliminate all sources of variation not related to the y -values. By using a wavelet analysis of the NIR spectra [52], the instrument-dependent information was removed from the y -dependent information. This enabled the development of a calibration model robust against instrumental differences. Finally, a more sophisticated way of preprocessing the spectra data is to use finite-impulse response filters [53], which eliminates the instrument-dependent information from the spectra.

The main advantage of this approach is that no a priori knowledge is necessary to develop a robust calibration model. This approach has been successfully applied to develop calibration models robust against instrumental fluctuations over time, as well as calibration models robust against instrumental differences between different spectrophotometers. However, each preprocessing technique can make the model robust only against a certain type of variation. For example, the multiplicative signal correction or the standard normal variate transformation can eliminate spectral differences due to particle size, but will not be able to correct very complex differences due to, for example, temperature variations. Therefore, this approach should be used with particular attention. For the moment, a number of promising preprocessing techniques have been successfully tested on a certain number of data sets. The limits of each preprocessing technique are not well defined. Therefore, one can never be sure that the selected preprocessing will give acceptable results in all cases.

11.4 INSTRUMENT RECALIBRATION

The last way of dealing with invalid predictions consists of selecting a subset of calibration samples that is representative of the original calibration data set. This subset should contain the spectral information relevant for calibration that is present in the original data set. When the current calibration model does not yield good predictions (fluctuations of the responses of a NIR instrument over time, differences between two NIR instruments, etc.), this subset can be remeasured using the new

measurement conditions and a new calibration model is determined with the new spectra of the subset samples.

In most of the encountered case studies, this approach is not satisfactory because one would like to avoid the recalibration procedure. However, there are some cases where this approach can be recommended. Firstly, all calibration samples should be perfectly stable over time, otherwise the new calibration model will account for both calibration information and sample aging. If this model is applied on new samples, erroneous predictions will be obtained since the new samples will not be properly located in the calibration domain because of sample aging. Another important point is the size of the recalibration subset. If the number of samples to be retained to cover all sources of variation is large, the use of the subset recalibration will be very time-consuming and the use of a smaller subset for standardization will be a better alternative. However, if the relevant spectral information can be easily summarized in a few calibration samples (yielding a parsimonious and simple model), this approach can be recommended, since the amount of work for the recalibration will be similar to the one for instrument standardization.

11.5 CONCLUSION

Transferring calibration models is important for every application, as all instruments will eventually fail. If preparations have been made beforehand, the transfer is facilitated. Without such preparations, the replacement of the calibration will usually involve as much work as the original calibration.

The types of calibration transfer algorithms differ substantially in their ease of use and applicability. Choosing the correct one can be difficult. However, a rule of thumb is that as the spectrophotometers become more similar, both the transfers are easier and the algorithm simpler. Some important points have been discussed in order to help users to select the most suitable approach. As for calibration, the complete transfer step should be carefully validated.

New algorithms and approaches are constantly being published. The final approach would be able to transfer calibrations within minutes and not be dependent on the condition, age, or type of instrument. This is a formidable task, and probably will not be reached soon, if ever. In the meantime, there will be improvements of a somewhat incremental nature, improvements that the user of transfers should be aware of in order to employ the best one.

Recently, the use of robust models has been proposed to obtain calibration models that are not sensitive to certain types of variations. Therefore, it enables one to reduce the number of situations where standardization is necessary. This approach is very interesting, because each standardization can be a delicate task. Both approaches can be very powerful and should be complementary. The development of robust models can strongly reduce the number of situations requiring calibration transfer, but standardization cannot be avoided in a number of cases where the differences are either unexpected or too complex.

REFERENCES

1. D. L. Wetzel, *Anal. Chem.*, **55**: 1165A (1983).
2. W. F. McClure, *Anal. Chem.*, **66**: 43A (1994).
3. J. Workman Jr., *J. Near Infrared Spectrosc.*, **1**: 221 (1993).
4. K. I. Hildrum, T. Isaksson, T. Naes, and A. Tandberg, *Near-Infrared Spectroscopy: Bridging the Gap between Data Analysis and NIR Applications*, Ellis Horwood, Chichester, 1992.
5. B. G. Osborne, T. Fearn, and P. H. Hindle, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, 2nd ed. Longman Scientific and Technical, Essex, 1993.
6. R. J. Dempsey, D. G. Davis, R. G. Buice Jr., and R. A. Lodder, *Appl. Spectrosc.*, **50**: 18A (1996).
7. P. K. Aldridge, R. F. Mushinsky, M. M. Andino, and C. L. Evans, *Appl. Spectrosc.*, **48**: 1272 (1994).
8. B. F. MacDonald and K. A. Prebble, *J. Pharm. Biomed. Anal.*, **11**: 1077 (1993).

9. M. A. Dempster, J. A. Jones, I. R. Last, B. F. MacDonald, and K. A. Prebble, *J. Pharm. Biomed. Anal.*, **11**: 1087 (1993).
10. W. Plugge and C. Van der Vlies, *J. Pharm. Biomed. Anal.*, **10**: 797 (1992).
11. C. Jones and J. A. Brown, *Adv. Instrum.*, **38**: 429 (1983).
12. U. Eschenauer, O. Henck, M. Huehne, P. Wu, I. Zegber, and H. W. Siesler, *Near-Infrared Spectroscopy: Bridging the Gap between Data Analysis and NIR Applications*, K. I. Hildrum, T. Isaksson, T. Naes, and A. Tandberg, eds., Ellis Horwood, Chichester, 1992.
13. A. F. Parisi, L. Nogueiras, and H. Prieto, *Anal. Chim. Acta*, **238**: 95 (1990).
14. M. S. Zetter and B. A. Politzer, *Hydrocarb. Process.*, **72**: 103 (1993).
15. P. Walling and J. Dabney, *J. Soc. Cosmet. Chem.*, **39**: 191 (1988).
16. S. Ghosh, *J. Text. Inst.*, **84**: 85 (1993).
17. P. C. Williams, K. H. Norris, and W. S. Zarowski, *Cereal Chem.*, **59**: 473 (1982).
18. O. E. De Noord, *Chemom. Intell. Lab. Syst.*, **25**: 85 (1994).
19. T. Dean and T. Isaksson, *NIR News*, **4**: 8 (1993).
20. T. Dean and T. Isaksson, *NIR News*, **4**: 14 (1993).
21. E. Bouveresse and D. L. Massart, *Vib. Spec.*, **11**: 3 (1996).
22. Y. Wang, D. J. Veltkamp, and B. R. Kowalski, *Anal. Chem.*, **63**: 2750 (1991).
23. R. W. Kennard and L. A. Stone, *Technometrics*, **11**: 137 (1969).
24. E. Bouveresse and D. L. Massart, *Chemom. Intell. Lab. Sys.*, **32**: 201 (1996).
25. P. Dardenne and R. Biston, *Proc. Third International Conference on Near Infrared Spectroscopy*, R. Biston and N. Bartiaux-Thill, eds., Agr. Res. Cent. Publ., Belgium, 1991, p. 655.
26. P. Dardenne, R. Biston, and G. Sinnaeve, *Near-Infrared Spectroscopy*, K. I. Hildrum, T. Isaksson, T. Naes, and A. Tandberg, eds., Ellis Horwood, Chichester, 1992, p. 453.
27. E. Bouveresse, D. L. Massart, and P. Dardenne, *Anal. Chim. Acta*, **297**: 405 (1994).
28. Y. Wang and B. R. Kowalski, *Appl. Spectrosc.*, **46**: 764 (1992).
29. E. Bouveresse, S. C. Rutan, Y. Vander Heyden, W. Penninckx, and D. L. Massart, *Anal. Chim. Acta*, **348**: 283 (1997).
30. E. Bouveresse, C. Casolino, and D. L. Massart, *Appl. Spectrosc.*, **52**: 604 (1998).
31. E. Bouveresse, D. L. Massart, and P. Dardenne, *Anal. Chem.*, **67**: 1381 (1995).
32. J. S. Shenk, *Proc. Third International Conference on Near-Infrared Spectroscopy*, R. Biston and N. Bartiaux-Thill, eds., Agric. Res. Centre Publ., Belgium, 1991, p. 649.
33. J. S. Shenk and M. O. Westerhaus, *NIR News*, **4**: 13 (1993).
34. J. S. Shenk and M. O. Westerhaus, U.S. Patent No. 4866644, Sept. 12, 1991.
35. Y. Wang, M. J. Lysaght, and B. R. Kowalski, *Anal. Chem.*, **64**: 562 (1992).
36. Y. Wang and B. R. Kowalski, *Anal. Chem.*, **65**: 1301 (1993).
37. Z. Wang, T. Dean, and B. R. Kowalski, *Anal. Chem.*, **67**: 2379 (1995).
38. B. G. Osborne and T. Fearn, *J. Food Technol.*, **18**: 453 (1983).
39. J. A. Jones, I. R. Last, B. F. MacDonald, and K. A. Prebble, *J. Pharm. Biomed. Anal.*, **11**: 1227 (1993).
40. N. B. Büchmann and S. Runfors, *J. Near Infrared Spectrosc.*, **3**: 35 (1995).
41. E. Bouveresse, C. Hartmann, D. L. Massart, I. R. Last, and K. A. Prebble, *Anal. Chem.*, **68**: 982 (1996).
42. B. Walczak, E. Bouveresse, and D. L. Massart, *Chemom. Intell. Lab. Syst.*, **36**: 41 (1997).
43. M. Forina, G. Drava, C. Armanino et al., *Chemom. Intell. Lab. Syst.*, **27**: 189 (1995).
44. M. Forina, C. Armanino, and R. Giangiacomo, *NIRS, in Near-Infrared Spectroscopy*, K. I. Hildrum, T. Isaksson, T. Naes, and A. Tandberg, eds., Ellis Horwood, Chichester, 1992, p. 91.
45. R. Goodacre and D. B. Kell, *Anal. Chim. Acta*, **348**: 511 (1997).
46. F. Despagne, B. Walczak, and D. L. Massart, *Appl. Spectrosc.*, **52**: 732 (1998).
47. O. E. De Noord, *Chemom. Intell. Lab. Syst.*, **23**: 65 (1994).
48. H. Mark and J. Workman, *Spectroscopy*, **11**: 28 (1988).
49. F. Despagne and D. L. Massart, *Anal. Chem.*, **69**: 3391 (1997).
50. S. Wold, H. Antti, F. Lindgren, and J. Öhman, *Chemom. Intell. Lab. Syst.*, **44**: 175 (1998).
51. J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, and S. Wold, *Chemom. Intell. Lab. Syst.*, **44**: 229 (1998).
52. J. Trygg and S. Wold, *Chemom. Intell. Lab. Syst.*, **42**: 209 (1998).
53. T. B. Blank, S. T. Sum, and S. D. Brown, *Anal. Chem.*, **68**: 2987 (1996).

12 Calibration and Validation of Process Sensors

Gary E. Ritchie

CONTENTS

12.1 Introduction	245
12.2 Sensors	247
12.3 Calibration and Validation.....	247
12.4 Validation of Process Sensors under the 21st Century Initiative	249
12.5 Risk Assessment, Risk Management, and Continuous Verification	251
12.6 Calibration of Process Sensors under the 21st Century Initiative	255
12.7 Future Considerations	257
12.8 Conclusion	258
Glossary of Process Terms Used in This Chapter.....	261
References	263

All models are wrong, but some are very useful.

Multivariate Calibration, Harald Martens and Tormod Næs, 1989

12.1 INTRODUCTION

In 1989, Harald Martens and Tormod Næs published their now seminal book, *Multivariate Calibration* (John Wiley & Sons Ltd, 1989). Although the book's primary focus is on what we now refer to as chemometric modeling — “soft modeling” to be more precise — they comment on an aspect of modeling that is at the present time very relevant to the focus of this chapter. To quote:

The optimal balance between theoretical model formulation and practical data analysis therefore depends on the extent of valid knowledge behind the model vs. the quality and relevance of the available data.

The same holds true for process sensors. It will be shown that, through transduction and transmission of signal into impulse, or through calculation into a determined value, the flow of process data into process information, which then may be transformed into process knowledge, can be turned into useful process understanding only if (a) the knowledge behind the sensor systems intended application can be shown to be valid; and (b) the quality of the data is accurately and precisely verifiable when that information is needed and is based on appropriate measuring reference systems. A further condition that is necessary is that the system must be adjustable (preferably in real or near-real time)

to offset changes in signal transduction timing or frequency and intensity, as information is fed back to the Supervisory Control and Data Acquisition (SCADA) system to induce a change to the process variable being affected.

Ambrozic and Parsons [1] imply that by using control charting and data summary tools, Real Time Quality Control (RTQC) made during batch evolution, corrections could be made to possible process upsets. They purport that by using adequate data from good batches, the mean of this data could then be used to establish " $\pm 3\sigma$ " limits on the process, which result in "6 sigma quality control" potential. However, what they did not touch upon, but obviously are aware of, is that none of this can be possible without having adequate sensors to measure the process variables of interest. In the context of this chapter, adequate sensors span the range of all available technologies that possess certain attributes and capabilities that define a sensor system.

In this chapter we address the issues stemming from calibration and validation of process sensors that will be used or interchanged with analyzers. Technically speaking, the establishment treats them rather differently. For calibration and validation issues, in most cases a sensor element may make up one component of the analyzer, and the overriding issue is where these instruments are going to be used (e.g., in the manufacturing suite). It is from this vantage point that we will treat sensors and sensor systems equally with analyzers.

The objective of this chapter will be to propose a scheme for demonstrating the accuracy, precision, and suitability of sensors. As will be seen, no matter what sensor, device, or measurement approach is chosen, these three simple criteria will be at the heart of the calibration/validation exercise. Depending on the complexity of the process, the sensor system, and process requirements necessary to achieve a mechanistic understanding, more parameters may be necessary. For example, linearity, range, specificity, robustness, ruggedness, detection limit, and quantification limit may also need to be investigated. By definition (see calibration and validation definition in this chapter), accuracy, precision, and suitability will be the minimum requirements for the calibration/validation exercise of process sensors.

Another aspect of process sensors or analyzers that should be self-evident when calibrating or validating those same instruments for off-line or laboratory use is that discrete measurements for accuracy and replicate measurements for precision under designed and controlled laboratory conditions for the purpose of achieving compliance to written instrument specifications should not be taken to constitute sufficient proof of performance assurance for those same instruments when placed in a manufacturing setting directly on-line or in-line to a process. The sensor and analyzer elements will interface with samples differently, the software elements will be used in a continuous, and not discrete mode, and a host of other differences simply too long to list here will impinge upon and affect operation of these instruments very differently than in the laboratory.

The quality assurance on instrument performance in the manufacturing suite has a completely different meaning in the context of calibration and validation exercises. What will be most critical to the assurance process will be to ensure that the quality of the signal and impulse as they affect and arise out of these instruments can be directly or indirectly correlated to changes in the processed material being measured. Having demonstrated that the correlation exists, the quality system will be continuously relied upon to check and perhaps adjust changes to its signal transforms or to impulse outputs in response to uncontrolled changes with assignable known causes arising from the physicochemical process attribute of the process material (i.e., signals have exceeded the upper and lower process set points). The decision as to when to have sensors and analyzers respond to process changes and to make adjustments, or to ignore and dismiss these signals (i.e., staying within the process set points) as controlled process changes without any assignable cause, is in fact the key decision for which one calibrates and validates.

This chapter explores the fundamental meaning and rationale behind measurement and control mechanisms and hopefully will lead the reader to view calibration and validation of process sensors as a very different exercise than the one that most analytical minds working in regulated industries have previously experienced in the laboratory.

12.2 SENSORS

A proper understanding of calibration and validation principles of process sensors must begin with explicit definitions of the terms. According to the National Materials Advisory Board (NMAB) report [2], a part of the National Research Council's (NRC) Division on Engineering and Physical Sciences (DEPS) under the auspices of the National Academy of Engineering (NAE), a single all-inclusive definition is not possible, given the various functional elements of "a sensor" and the continuous and rapid evolution at the physical, mechanical, electrical, and optical interfaces. Therefore, the committee adopted the following definitions:

Sensor element: The fundamental transduction mechanism (e.g., a material) that converts one form of energy into another. Some sensors may incorporate more than one sensor element (e.g., a compound sensor).

Sensor: A sensor element including its physical packaging and external connections (e.g., electrical or optical).

Sensor system: A sensor and its assorted signal processing hardware (analog or digital) with the processing either in or on the same package or discrete from the sensor itself.

The preceding definitions describe what are commonly referred to as *hard sensors*. According to Merriam-Webster's Online Dictionary, a sensor is a device that responds to a physical stimulus (as heat, light, sound, pressure, magnetism, or a particular motion) *and* (italics mine) transmits a resulting impulse (as for measurement or operating a control). Therefore, a second function of a sensor emerges from this definition, and it is the one most critical to this discussion, that being the transmission of a resulting impulse. Impulses can be used for a multitude of process activities for the purpose of controlling processes, for instance through SCADA. For the purposes of this chapter, sensors transduce a signal and transmit an impulse. Therefore, the term measurement should be considered as *both* the transduction *and* transmission of signal and impulse, respectively.

Another very important category of sensors is referred to as *soft or intelligent sensors*. Soft sensors are defined as an algorithm used to calculate a value by measuring other values [3]. The underlying principle of operation of soft or intelligent sensors is that they do not measure process variables directly, but they derive values from the impulses by calculating relationships among those impulses. Accordingly, a succinct definition for sensor for the purposes of this chapter is as follows:

INSIGHT 1



A sensor is any device that transduces a signal, and transmits an impulse, or measures and calculates a value.

Having defined sensors, a discussion of the current Good Manufacturing Practice (cGMP) definitions for calibration and validation follows with a presentation for a proposal for a new model for validation practices within the scope of the Food and Drug Administration 21st Century cGMP Initiative [4].

12.3 CALIBRATION AND VALIDATION

Calibration and validation practices in pharmaceutical manufacturing facilities are currently being driven by cGMP requirements [5,6]. The objective for cGMP is to ensure that pharmaceutical manufacturers can demonstrate unequivocally that their product has not been adulterated if tested.

An adulterated product is one in which a product is unfit to be sold as a result of

- Contamination
- Mix-ups
- Errors
- Having been manufactured under conditions which were not in a state of control.

Under these practices, the objective — calibrating and validating a process sensor element, sensor, and sensor system used in a pharmaceutical manufacturing facility — is to demonstrate unequivocally that a process can be measured for the absence of contamination, mix-ups, and errors. When are cGMPs applicable to process sensors? For the purposes of this discussion, cGMPs are applicable to process sensors whenever the manufacturer “... monitor[s] the output and ... validate[s] the performance of those manufacturing processes that may be responsible for causing variability in the characteristics of in-process material and the drug product (21 CFR Part 211.110).”

One might wrongly conclude, however, that achieving these criteria constitutes a demonstration that a process was in a state of control. As we will see later on in the chapter, calibration and validation of process sensors only verifies that the measurement system is suitable for its intended purpose, that being that it has been shown to accurately and precisely transduce the signal for which it was calibrated and validated, and the impulse is transmitted accurately and precisely to its intended target. Much more work would be required to also demonstrate that the measurement indicates that the process (i.e., material or final product) also has or has not been adulterated. This is a very important distinction, and must be kept in mind as one seeks to employ sensors to obtain process understanding. The use of valid sensors for concluding that a “clean process” has been run is not tantamount to attaining process understanding unless a relationship has been previously demonstrated to exist between the measurement and the process variable of interest. To infer causality from apparently “valid” sensors without first having demonstrated correlations or equivalence to physicochemical changes in process material is nothing short of flirting with disaster.

This is because, as we shall see later on in this chapter, process variability can be shown to arise primarily from two sources: both controlled and uncontrolled variation. Controlled variation does not have any known assignable cause arising from variation of the process or process material, whereas uncontrolled variation does. What must be ultimately discerned is which sources of variation can be shown to be the result of assignable causes that can lead to potential adulteration of the process and hence product. Keeping this in mind, we now can define what calibration and validation cGMPs are.

Current GMP practices define calibration and validation as follows:

Calibration: The process of periodically checking and adjusting measuring devices and instruments to ensure the specified accuracy and precision that are traceable to national or international standards.

Validation: The establishment of documented evidence that a system does as it purports to do.

Under these practices then, calibration and validation of process sensor elements, sensors, and sensor systems used in a pharmaceutical manufacturing facility would have to be

1. Periodically checked
2. Adjusted to maintain accurate and precise measurements
3. Traceable to national or international standards

and all of this would have to be documented.

Under the new 21st Century cGMP Initiative, the meaning of the above cGMP requirements would essentially remain unchanged with respect to a pharmaceutical manufacturer having to be


able to demonstrate unequivocally that its product has been shown to be unadulterated if tested. However, under one of the new guidances of the initiative, specifically the Process Analytical Technology (PAT) Guidance [7], the approaches taken in order to achieve calibration and validation requirements will need to be addressed. Essentially, none of the three requirements listed above can be achieved in practice under the new initiative as written. The new initiative will require that

1. The frequency of calibration requirement be addressed.
2. The in-process adjustment of sensors to maintain their functional state in order that transduction, transmission, and measurement and calculations be achieved in real or near-real time must be addressed.
3. The use of traceable standards has to be reconsidered in light of current and proposed capabilities and future requirements.

all within a framework of risk, data assessment, and management approaches with the sole objective of ensuring that safety, purity, and efficacy of a final product are not compromised by material, process, or sensor failure.

The elements of calibration and validation of sensor systems are thus composed of all of the physical attributes that make up sensor systems, the functional aspects that determine the operational capability of these systems, and finally the parameters that determine the frequency of verification, the sensor system variables that determine the operational range, and the process/sensor standards that will be used during the verification phases of the sensor systems.

INSIGHT 2

 *Calibration of process sensors is verified at minimum by demonstrating accuracy and precision of signal transduction, and that the resultant impulse is transmitted accurately and precisely to its intended target.*

What follows are schemes for a proposed approach to consider for calibration and validation of processes within the framework of the PAT guidance.

12.4 VALIDATION OF PROCESS SENSORS UNDER THE 21ST CENTURY INITIATIVE

Figure 12.1 is a diagram of the principles of validation within the scope of the 21st Century Initiative. Sound knowledge of calibration and validation approaches for process sensors can be shown to be the key on which all of PAT is based on. Essentially, true process understanding will best be demonstrated by those capable of being able to ameliorate calibration and validation issues by achieving continuous verification of measurement systems under variable conditions. Proper alignment and implementation of PAT, process control technology (PCT), and process assessment technology (PAsT) tools with the right process critical control parameters (PCCPs) adds up to obtaining a fundamental understanding about measurement performance of the process. Achieving process capability through risk assessment, risk management, and continuous improvement of the process closes the process understanding loop. These result in a fundamental understanding about management performance of the process. Together, performance measurement and performance management result in a high level of sustainable process understanding through a calculated and deliberate iterative process.

From these essential underpinnings, process data can be gleaned from PCCPs that can lead to real-time or near-real-time process information, which can be processed into process knowledge to arrive at process understanding. Having achieved understanding, the decisions made within the

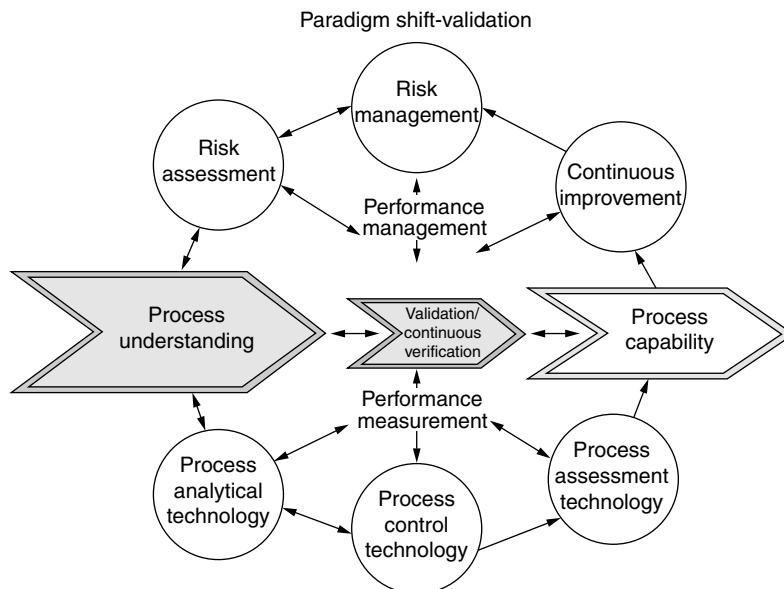


FIGURE 12.1 PAT process validation. (Courtesy of Ken Leiper.)

process feedback or feed-forward loops, automatically, semi-automatically, or by human intervention mode, should always result in the desired state of consistent and predictable quality target attributes for the finished product.

Cohen [8] presents an historical perspective of validation practices as they were instituted and evolved from laboratory testing to on-line testing of resistivity/conductivity and total organic carbon (TOC). A review of that paper shows that validation before PAT required the following validation tests for an on-line water system:

1. BPE guidelines
2. ASME specifications
3. ISPE Baseline[®] Guide on Water and Steam Systems
4. Installation qualification (IQ)
5. Operational qualification (OQ)
6. Performance qualification (PQ)
7. SOPs
8. Maintenance procedures
9. Corrective action plans
10. Internal corporate documentation

These documents were all designed to support an audit of the system to answer questions about design, installation, initial operation, validation, and continued operations of a water system. However, since that time better and more appropriate methods have been made available to accomplish the same tasks, but in a more efficient manner. Cohen illustrates the use of electronic documentation approaches for accomplishing the same validation tasks, but in real-time. The improper use of the term “revalidation” limited the use of these real-time approaches, when in fact, as Cohen points out, what was intended was “change control.” In reality, what was meant to be accomplished was not revalidation of the water system, but validating “change controls.”

Ritchie and Hashemi [9] demonstrated a validation approach for an on-line near-infrared (NIR) spectrophotometer that used a computer installed on a network. Their validation approach for the

computer-controlled device embraced the current concepts of PAT. They suggested that the design specifications (and the functional requirements) should spell out in detail how software and hardware are integrated to be able to perform six basic functions:

1. Collection
2. Calculation
3. Control
4. Archiving
5. Retrieval
6. Restore

Furthermore, they implemented change control requirements that fulfill Cohen's observation for validating "change control" and not a "revalidation" of the total system.

12.5 RISK ASSESSMENT, RISK MANAGEMENT, AND CONTINUOUS VERIFICATION

The PAT guidance recommends that risk-based approaches be applied to PAT systems. According to the Society of Risk Assessment (www.sra.com), risk is defined as:

1. The potential for realization of unwanted, adverse consequences to human life, health, property, or the environment; estimation of risk is usually based on the expected value of the conditional probability of the event occurring times the consequence of the event given that it has occurred.
2. Thomas Cool provides an alternative definition of risk in the context of uncertainty [10].

Much more could be said about risk, and there are many resources both in the public and private sector that deal with this topic in depth. However, the important issue that must be focused upon when considering using risk-based approaches in the context of PAT is that the intent of the guidance is to provide for the ability of a manufacturer possessing a high degree of process understanding, and a well-developed and capable quality system for that process, to use less restrictive regulatory cGMP approaches to manage change (see discussion in section on Validation of Process Sensors under the 21st Century Initiative in this chapter).

The emphasis on uncertainty is perhaps more telling for applying the concept of risk to PAT, and as shall be demonstrated, is very much in line with what is intended in monitoring PCCPs with sensors that are distributed throughout a process via SCADA systems that utilize multivariate statistical process control (MSPC) methodologies for controlling these processes.

In the context of the FDA PAT guidance, risk analysis is assessment, management, and communication of an approved products process. So for a process that is purported to be fully understood by a drug manufacturing firm, the Agency seeks to provide assurance to the public that an approved product meets its expectation that a firm can demonstrate sound scientific, and risk-based knowledge about the products process, that it can demonstrate sound risk-management practices, and that a firm has effectively communicated that understanding to the Agency.

However, risk in the process world has an even more fundamental role that is far more just to fulfill the Agency's expectation about process understanding and product safety: it is the basis and rational for MSPC. In fact, it is what the original intent of Walter Andrew Shewhart (1891–1967) had in mind when he invented the notion of statistical process control (SPC) and the control "Shewhart" chart. Although not couched in precisely the language of risk, it was at the heart of what he was trying to do at Bell labs at the time [11].

Neave and Wheeler [12] present an excellent discussion on the proper use of control charts and show how the notion of probability introduced into the application by statisticians since its inception by Shewhart has actually promulgated false notions about Shewhart's original intent when he stated, "We are not concerned with the functional form of the universe, but merely with the assumption that a universe exists."

Neave and Wheeler maintain that the use of risk and the notion of the probability of risk as the rational behind SPC and classically *how* risk is calculated to yield assumptions based on the calculated probability that something risky is about to occur has been wrongly attributed to Shewhart's thesis on SPC. They maintain that this is not what Shewhart meant to imply, and argue that Shewhart actually goes to great lengths to argue against the use of this kind of statistical approach for SPC, because it is not necessary.

This paper is so important in order to be able to understand the underlying intent of the Agency and their PAT guidance, and *how* they want PAT to be applied, that some of its content bears repeating here. This work can and should become a source from which calibration and validation practices of process sensors and analyzers are formulated as well. Once it can be established that a meaningful change (rather than a random change) can be correlated with an assignable cause, calibration and validation of the sensor for its measurement potential, and that potential is related to a PCCP that has a high level of certainty of being related to the product quality attribute of interest, then process understanding can be said to have been achieved.

It may also be true that if it can be demonstrated that changes in PCCPs that result in a consistent and predictable "adulterated" process material or finished product, and the correlations from measurement to PCCP and from PCCP to a product attribute can be achieved with a high degree of certainty, then this too should be a valid demonstration of process understanding. Having achieved the latter is not a bad thing; in fact, it is perhaps even more desirable actually, because the process controls needed to be assessed and validated can be checked against an assignable cause known to have a negative effect on the final product quality attribute. Restated, both positive and negative measurements on PCCPs that are directly, inversely, or inferentially related to changes in process composition, that are shown to be related to the desired final product quality attribute, are all equally valid measurements for demonstrating process understanding.

In Neave and Wheeler's Shewhart's Charts and the Probability Approach, they suggest the following key points about SPC must be strictly understood and adhered to:

1. The basic concept behind Shewhart's control chart is the distinction between two categories of variation:
 - a. "controlled variation."
 - b. "uncontrolled variation."
2. Controlled variation, while appearing to be predictable at first (in that you can only predict "something" is changing), does not have an assignable cause.
3. Uncontrolled variation can be shown to be the result of an assignable cause, and physicochemical changes in the process material can be monitored for that occurrence.
4. The control chart is a technique for detecting which type of variation (controlled or uncontrolled) is being displayed by a given process.
5. The objective of the control chart is to give the user a guide for taking appropriate action to look for assignable causes when the data display uncontrolled variation, and to avoid looking for assignable causes when the data display controlled variation.
6. Since the decision described above is an inductive inference, it will always involve some degree of uncertainty.
7. One of two mistakes can occur when decisions of this sort are made:
 - a. The first mistake is to attribute an outcome to an assignable cause when it is simply the result of common causes.

- b. The second mistake is to attribute an outcome to common causes when it is, in truth, the result of an assignable cause.
- 8. It is impossible to avoid both of these mistakes, so this cannot be the objective of SPC.
- 9. Instead, a realistic objective is to minimize the overall economic loss from making these two mistakes.
- 10. To this end, Shewhart created the control chart with three-sigma limits.
- 11. We are not trying to find some exact model to describe the process, but we are trying to determine if the process fits (at least approximately) a very broad model of “random behavior.”

Point 11 is counter-intuitive to what one normally thinks about modeling processes, but that is because Shewhart’s approach was not founded within the context of process monitoring. It was founded on the premise that, given all processes undergo change, that change, used as indicators of root causes, can be used to *improve* the process. This is indeed a profound leap of thought that Shewhart attained, but it is the point that Neave and Wheeler show has been most misrepresented by statisticians. And this is also precisely the rationale given in the PAT guidance: “Effective use of the most current pharmaceutical science and engineering principles and knowledge — throughout the life cycle of a product — can *improve the efficiencies of both the manufacturing and regulatory processes. This FDA initiative is designed to do just that by using an integrated systems approach to regulating pharmaceutical product quality.*”

Keeping the pharmaceutical manufacturing industry focused on this approach from the outset of the design, qualification, installation, and during continuous operation and maintenance of their processes will ensure that successive manufacturing campaigns using a particular validated process can be continuously verified for meeting the in-process material specification (continuous verification and validation). Another way to interpret what this means is to restate it in the following way:

Continuous verification and validation means that data from the measurement is checked against the specifications as it is being acquired (in real or near-real time). The system and data are said to be “in control” and “valid,” respectively, when the specifications are not exceeded (the process data values are not less than or greater than the lower and upper specifications limits previously established in the design and qualifications stage of process development). Neave and Wheeler take the point even further by suggesting what Shewhart’s intent was for plotting process data:

One can use the chart to understand how a process is behaving. Therefore, a process is said to be “in control” only when it has demonstrated, by means of a reasonable amount of data, that it is performing consistently and predictably. Moreover, from a practical perspective, such a process is, in essence, performing as consistently as possible. Therefore, any process which is not performing “as consistently as possible” may be said to be “out of control.” Thus, the control chart is a procedure for characterizing the behavior of a process.

Repetition of this measurement step each time a new campaign is initiated is tracked and assessed for variance in the distribution, trends, and outliers in the data measurement and risk (continuous fault) analysis on the sensors or analyzers and CPUs. All or part of the measurement and sensor data may become part of a database constructed from past validation and historical data that have been previously shown through root cause analysis to be the result of *known* material, process, or sensor variation that has exceeded the specifications. Each manufacturing campaign becomes a potential well for new data, which in turn becomes fodder for a dynamic database for which variable data can be assessed against nominal and expected process and sensor performance, diagnosis made, and remedial action instituted, all occurring perhaps in a millisecond to a second timescale.

The risk assessment work done up front during the design, qualification, and installation stages now come into play when these specifications are exceeded. Exceeding the process specification limits would not necessarily imply a nonvalid system, but are the basis for establishing process

alarms that serve as the trigger for real or near-real time corrective measures. Either a feedback or feed-forward correction would need to be made to bring the system back to its process set points. In this sense, risk assessment is not something that is studied and used just during the design and qualification stages, but becomes a continuous function, a dynamic function that may be used hundreds, even thousands, of times a second during the routine course of the process for which it was designed and intended.


Now what takes place next strikes at the heart of the present system and is the essence of the proposed PAT approach. Suppose a batch of incoming raw material tests low or high upon receipt, or the certificate of analysis (CoA) accompanying the material states that a certain “new” or additional test gives a value for a physical parameter that has a known effect on a particular step in the process. Based on “process understanding” and a highly integrated quality system in place, a set point is potentially subject to be changed. The risk for doing so with very little process understanding would obviously be high. And a quality system, under current cGMP criteria, and charged with preventing this kind of tampering with the process, would simply not allow a change like this to be made, not without a thorough and extensively documented investigation prior to either granting approval for changing the set point or rejecting the change because it is not allowed without being able to show assignable cause for the change.

However, under the PAT approach, a process designed with quality and risk-based practices built in from the inception right on through the launch of the first production batch will not have any problem accommodating a change such as is being illustrated here. Based on experience with the overall factors that contribute to the variability, predictably based on prior knowledge of how the process system (equipment, sensor, etc.) behaves, and from the variability inherent in the process materials, this risk-based change is the moving target, if you will, of a well-understood process. It is precisely what the process should be designed for and what the calibration and validation of process sensors and analyzers should be based upon.

The preceding scenario is possible only if a well-planned, rigorous, and seamless deployment scheme is adopted early on. The fault of not being able to think about fully integrated systems that must include risk assessment and management tools throughout the sensor systems, as well as other tools for designing, qualifying, installing, implementing, measuring, controlling, and maintaining these processes, stems from linear-type thinking. Event sequencing may be useful for arriving at how one might approach solving a problem, but then it is almost never how one would proceed to actually put it into practice. Current thinking about PAT tools is a case in point. The risk-based tools used to design the process and to qualify the systems can be used to calculate the probability that, given that the correct PCCP has been correctly identified and validated, they can also be continuously implemented to calculate the probability that a failure event may be about to occur. This is a completely different use of risk assessment from just design and qualification, but this is precisely what is required to have full process understanding. How is that? It is possible to prove that causality of sensor measurement of PCCPs can be demonstrated by not just being able to show what process variables or process material variables are responsible for trending or outliers, but also that risk calculations based on these variance deviations can and should be used to trigger corrective actions by affecting sensor systems and other components (i.e., CPU's) in the process train. Of course, this assumes that the process steps affected are reversible, and can be stopped and restarted, slowed, or at least can be completely halted before major malfunctions occur that knowingly could lead to adulteration of product during processing. This is the intent behind SPC charting and what PAT is meant to accomplish.

Risk-based approaches should and can be used at major points along the process train. From the sensor system to the PCCP and the process material itself (perhaps based on ancillary classical process or laboratory measurements, such as temperature, pressure, humidity, pH, viscosity, particle size, hardness, etc.) and any other process variable deemed to add value to the overall process understanding, all these can be used to mine data (i.e., build databases) that can be used to extract useful process information on which real-time or near-real time decisions can be made.

INSIGHT 3

 *Validation of process sensors is verified by demonstrating the suitability of the measurement to some uncontrolled variation shown to be the result of an assignable cause arising from physicochemical changes in the process material. SPC and chemometric, multivariate analytical approaches can be applied to cause the sensor to direct the process critical control parameter (PCCP) into a “satisfactory state” of minimum variance.*

12.6 CALIBRATION OF PROCESS SENSORS UNDER THE 21ST CENTURY INITIATIVE

Keeping in mind the nominal requirements for cGMP compliance requirements for lab-based instruments for demonstrating calibration (checking, adjustments, accuracy, precision, and traceability), what are some of the capabilities that will need to exist for process sensors or monitoring systems, and how will they be calibrated?

Given that the premises for cGMP requirements are all now being challenged, and even newer capabilities are being added to sensors, it would be wise to begin with a description of what capabilities are being demanded for current sensor and monitor systems. A current sensor or monitor system should be able to meet the following requirements:

1. Automated, remote control, and instantaneous verification of key sensor elements at any time
2. Programmable or automated self-adjustment of key sensor elements
3. On-demand instantaneous read out of key sensor elements statistics
4. Embedded calibration elements for self-referencing capability.

All of the capabilities listed above for the most part can be categorically put under the moniker of “smart” sensors. Further enhancements being demanded by the market are wireless, compact, small or even miniaturized, durable, inexpensive, and, of course, fast. Obviously, the more capabilities these sensor systems possess, the more control one can exert over a process; however, a price must be paid for every feature. That price is increased assurance.

While it may be true that enhanced and consistent product quality can be achieved by enhanced process understanding, the price to achieve that can be directly correlated to the increased cost and effort spent to assure that the sensor system, being the source from which process understanding is derived, is of equal or better quality. Quality then, being the operative term, is the parameter that should be the focus of study if one seeks to improve process capability through the use of enhanced sensor or monitor systems. The other variable that can and should be studied is the time spent currently measuring the product vs. the time spent in the future “monitoring the monitors” (sensor systems).

Quality aside, a pragmatic approach is needed to bring together the elements of design, qualification, implementation, and real-time quality assurance (RTQA) of these new concepts. Figure 12.2 brings together these concepts, showing how each piece of the requirements previously discussed fit in with one another.

Much has been written about “Quality Variables” in the context of PAT, on how they are identified, selected by design of experiment (DOE), and established as primary causative agents for final product quality outcomes. A closer look at what DOE actually does can lend insight into what characteristics to look for in-process data that can determine how to identify, select, and establish the principal quality variables from a process for which a sensor would then need to be calibrated. Using this

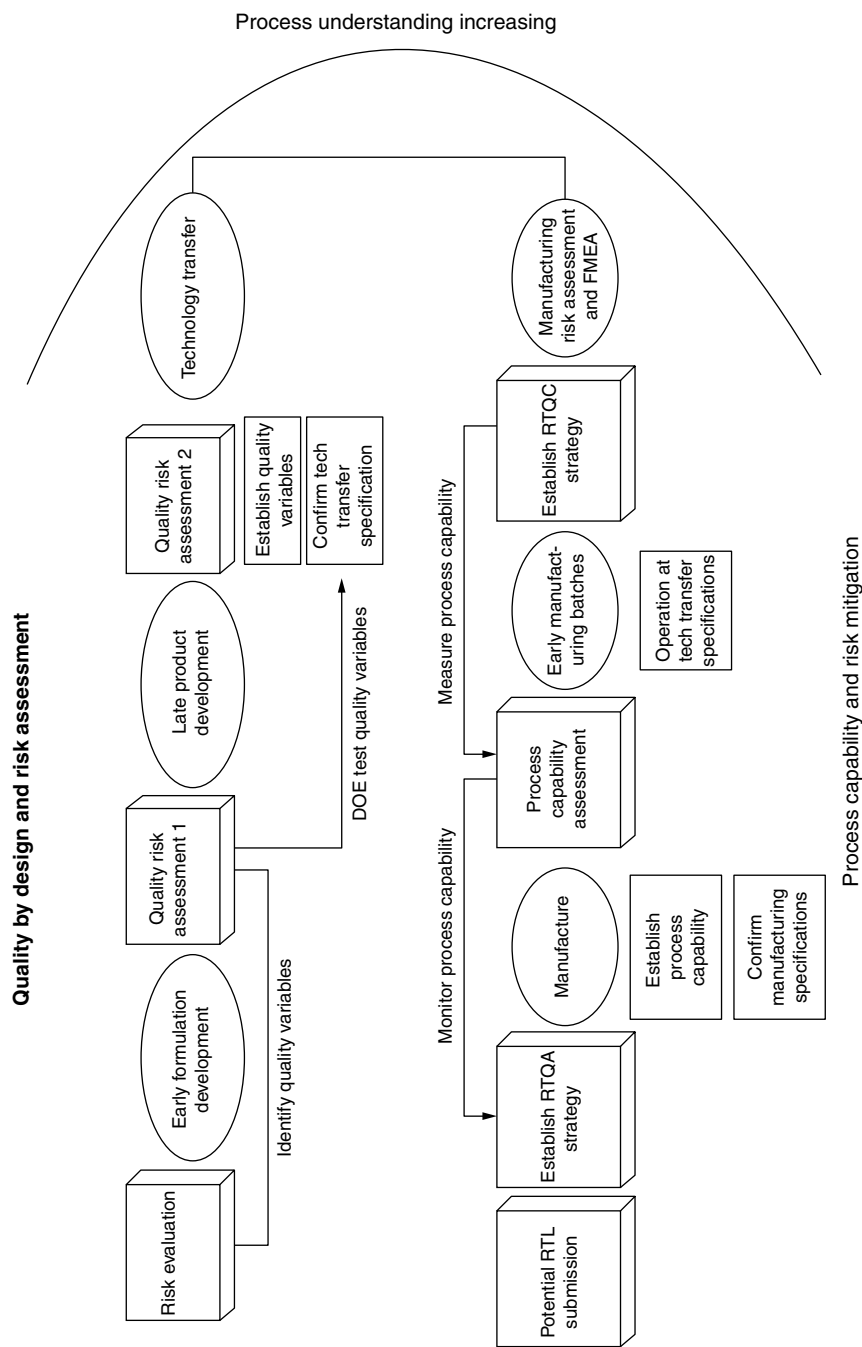


FIGURE 12.2 PAT process map. (Courtesy Robert S. Chisholm.)

approach, rather than following a prescriptive regulation that requires one to calibrate a sensor system for everything, the objective would be shifted to only the primary variable or variables giving rise to the process signal for which one is calibrating in the first place.

It has been the practice to adopt instrument performance verification requirements to the range of capabilities that exist among the hardware/software configurations. With the advent of PAT, this can no longer be a tenable approach. As more and more knowledge is gained about just exactly how pharmaceutical processes operate, from API synthesis and excipient raw material — API interactions — to the final finished product, more and more instrumentation is going to be made to exacting specifications that will require a different set of approaches from the prescriptive one currently being applied. For instance, if a particular sensor system is to be calibrated for a quality variable that is to be measured every tenth of a second, does calibrating that sensor for linearity have any meaning?

Precision, range, and accuracy would be more appropriate. And if it is a requirement to know precisely the time interval of the course of a particular sequence of chemical reactions during a process (e.g., because of known chemical kinetics and stoichiometry using that particular sensor), then perhaps calibrating the sensor against a clock to determine how accurately and precisely it can measure intervals (e.g., every tenth of a second by the use of an on-board clock traceable to the U.S. Naval Observatory Atomic Clock) would be even more important than accuracy. For process understanding to be realized, a major factor has to be the realization that those regulations for demonstrating sensor suitability — in order to have meaning in the context of *timely and relevant* real-time or near-real-time process measurements — should be achievable in a parsimonious way, without being encroached upon by artificial, arbitrary, and unnecessary requirements.

12.7 FUTURE CONSIDERATIONS

The next generation of sensors being developed for use in pharmaceutical manufacturing production facilities that will achieve parsimony can be found at the University of Washington (Seattle) Center for Process Analytical Chemistry (CPAC). The New Sampling/Sensor Initiative (NeSSI) is an attempt to reduce design, construction, installation, operation, and maintenance costs of manufacturing processing and analytical systems across multiple industries. NeSSI was initiated by the Instrumentation, Systems, and Automation Society (ISA) in 2000.

The premise of the CPAC NeSSI can be gleaned from the following outline [13]:

Facilitate the acceptance/implementation of:

- Modular, miniature, and smart sample system technology
- Based on semiconductor (SEMI) standards & ISA SP76 work
- Adapt for chemical industry (traditional modular).

Promote the concepts of:

- Field-mounted analytical systems
- Integration of sample systems with physical/chemical sensors (Plug and Play)
- Small, smart, integrated sampling, sensor, and analyzer transmitters
- Validation of representative sample and analysis.

Lay the groundwork for the next generation of Process Analytical Systems:

- Open connectivity architecture (Ethernet, TCP/IP, wireless etc.)
- Industry standard protocols (OPC, Fieldbus, Profibus, Hart, etc.)
- Web-enabled technology (browsers, HMI, I/O servers, etc.).

Longer Term:

- Modular systems provide bridge between the macro world of process streams and the micro world of new generation of microfluidic/analytical systems (e.g., analyzers on a chip, sensors on a chip, lab on a chip, etc.).

The NeSSI approach could potentially provide opportunity for achieving the desired state that is being defined by the FDA PAT initiative. From the point of view of how calibration objectives could be met, the PAT guidance indicates that “Strategies should accommodate the attributes of input materials, *the ability and reliability of process analyzers to measure critical attributes*, and the achievement of pre-established process end points to ensure consistent quality of the output materials and the final product.”

Some of the NeSSI approaches and criteria being discussed for achieving a high degree of assurance that process sensors will function reliably are

1. Transducer electronic data sheets (TEDS) architecture — calibration TEDS — storage of embedded calibration parameters
2. Remote calibration and diagnostics architecture
3. Device type manager (DTM) controllers architecture
4. Sensor actuator manager (SAM) controllers architecture.

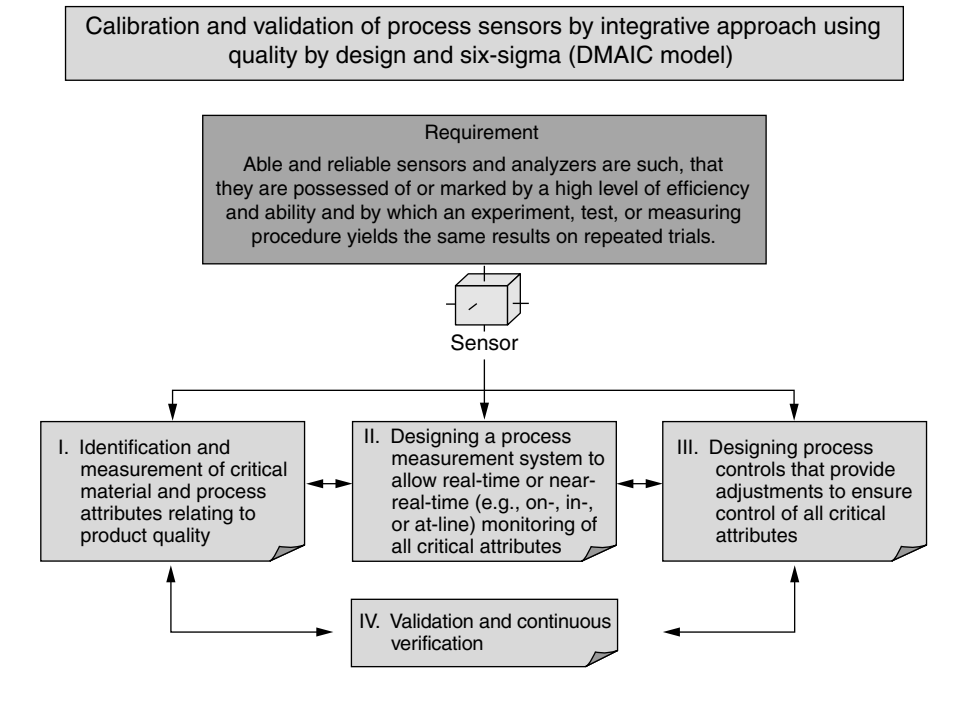
All of these approaches attempt, in one way or another, to achieve reliable calibration strategies. By reducing the cost, by either use of redundant embedding of pertinent calibration programs for optimizing and maintaining sensor performance (i.e., use of micro electro mechanical’s [MEM] devices) or via remote, wireless networking architecture (i.e., WIFI, satellite communication), both approaches represent radical departure from current instrumental designs. The wireless and satellite approaches, while perhaps drastically more expensive to implement, offer the greatest potential for multiple, wide area, and global networking capability.

12.8 CONCLUSION

The present climate surrounding the pharmaceutical manufacturing environment is rapidly undergoing change due to various internal business pressures and external forces that extend beyond the business realm. The focus of the regulatory agencies on ensuring overall product quality has shifted from a product-centered quality assurance model to a process-centered model. In doing so, cGMP requirements, while still wedded to the premise “that a pharmaceutical manufacturer can demonstrate unequivocally that his product has not been adulterated if tested,” should be interpreted differently with respect to how a firm executes these cGMPs in a process-centered quality-assurance environment and how the regulatory authorities will enforce them.

With respect to cGMPs that will be used to regulate how process sensors will be calibrated and validated, new application and architectural schemes are being proposed for the design, qualification, installation, operation, and maintenance of in-line and on-line sensors.

From the preceding discussions on sensors, calibration and validation cGMPs, and risk, a new view has emerged on management of sensors used for real or near-real-time analysis in the pharmaceutical manufacturing process. Figure 12.3 represents how the PAT approach for cGMPs for the 21st Century impacts how sensors will be calibrated and validated by a quality-by-design (QbD) approach for ensuring that final product quality attributes are achievable with little (based on Six Sigma 3.4 defects per million opportunities, DPMO) or zero failures (based on Total Quality Management [TQM] practices of zero-defect). An additional approach to be considered for meeting the goal of a “Six-Sigma” pharmaceutical manufacturing process will be the principle heralded in the



Calibration and validation	Risk	QbD	Six-sigma
I	What is it I want to measure? Why do I want to measure it? Where will I measure it? When will I measure it? How will I measure it?	DOE	Define opportunities
II	Relate PCCP to measurement result to product quality attribute	Verify sensor and analyzer ability by MSPC	Measure performance
III	What is it I want to control? Why do I want to control it? Where will I control it? When will I control it? How will I control it?	Experiment, test, or measuring procedure	Analyze opportunities
IV	Achieve consistent and predictable product quality attribute	Reiterate back to steps I, II, or III	Improve performance control performance

FIGURE 12.3 Strategy for demonstrating ability and reliability of on/in-line process sensors and analyzers.

Six-Sigma approach of DMAIC [14]. The DMAIC model includes five phases:

1. Define opportunities
2. Measure performance
3. Analyze opportunity

4. Improve performance
5. Control performance.

A process that is dependent on the performance of sensors will require a thorough assessment of quality programs linked to the process. Even more important than this will be the fact that quality data obtained from these systems must be used to make critical decisions. Some of these decisions may be automatic, semi-automatic, or require human intervention to activate the next steps in the process. All of these decisions will be highly dependent on the first steps conceived at the design stage. Perhaps the most critical decisions that will ever be made will be the choice of the correct answers to the following questions:

1. What is it I want to measure?
2. Why do I want to measure it?
3. Where will I measure it?
4. When will I measure it?
5. How will I measure it?

Put in the context of calibration and validation requirements, the answers to the four W's and H, and their objective of achieving an able and reliable sensor that can be depended on to also control a process, will be needed. As it has been stated in the PAT guideline, "Design and construction of the process equipment, the analyzer, and their interface are critical to ensuring that collected data are relevant and representative of process and product attributes. Robust design, reliability, and ease of operation are important considerations." With respect to reliability, which in this context means assurance as demonstrated by way of calibration and validation tests performed on sensors or analyzers, consideration should be given to:

- Identification and measurement of critical material and process attributes relating to product quality
- Designing a process measurement system to allow real-time or near-real-time (e.g., on-, in-, or at-line) monitoring of all critical attributes
- Designing process controls that provide adjustments to ensure control of all critical attributes.

Validation and continuous verification of a quality process by definition will be highly dependent on the experimentally determined outcomes for assessing the ability and reliability of the sensor systems and analyzer technologies employed to monitor, control, and analyze the process material composition. As can be seen from Figure 12.3, calibration and validation steps are integrated throughout the four primary guidance principles, which are the drivers for the QbD and Six-Sigma steps. This is radically different than the calibration and validation approaches traditionally undertaken in the laboratory. Figure 12.4 [15] illustrates traditional calibration and validation approaches taken in the laboratory. As one can see, installation qualification (IQ)/operational qualification (OQ)/performance qualification (PQ) done at intervals determined by installation, qualification and during normal wear and tear for change control purposes, do not answer the imperatives presented by the four primary guidance principles (see I to IV in Figure 12.3). Table 12.1 compares calibration and validation requirements for laboratory-centered instrument quality vs. process-centered instrument quality.

Kemeny [16] lists some requirements of process analyzers that are also relevant to sensor systems as well, and is worthwhile repeating, with some slight modifications, here. These simple, uncomplicated yet elegant truths are basic to meeting success as pharmaceutical manufacturers expand and explore their manufacturing capabilities in the 21st Century:

- Process analyzers are integral parts of the process loop. Therefore, their parameters have to match the requirements of the control loop of the particular process.

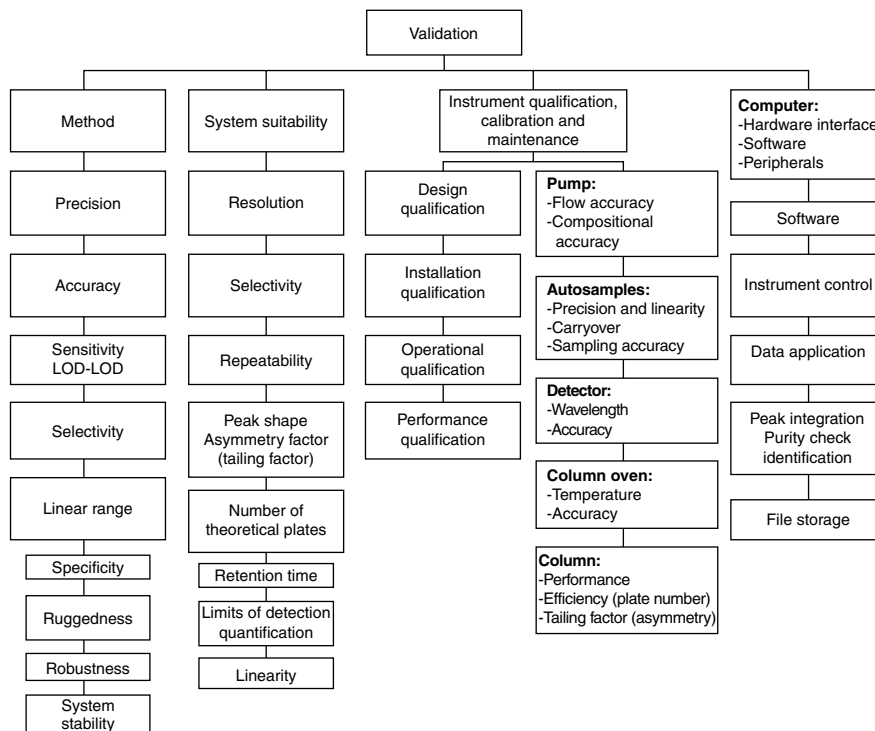


FIGURE 12.4 Traditional calibration and validation strategy for a laboratory-based high performance liquid chromatography (HPLC). (Reprinted with permission from Marcel Dekker Inc.)

- Continuous production processes are designed for 24-h operation. Therefore analyzers should also be capable of continuous uninterrupted operation.
- Analyzers should be reliable enough to provide analytical results over long periods of time.
- Analyzer failure should be able to be detected by built-in diagnostic routines to detect faulty conditions.
- Failure should be accompanied by error message.
- Error messages could be in the form of a digital message (fully captured and archived), or a simple digital logic signal connected to the process controller receiving the analyzer output.
- Error messages must be categorized as hardware, software, or process errors.
- The only planned interruptions in the operation of the analyzer should be the calibration/recalibration and the maintenance.
- The enclosure designs have sound well-tested practices. To support this, there are design testing guidelines, standards, testing companies, agencies, and consultants.
- The purpose of the analysis is to use the data to control the process.

GLOSSARY OF PROCESS TERMS USED IN THIS CHAPTER

CPU: Computer processing unit.

PCCP: Process critical control parameter.

RTQC: Real time quality control.

TABLE 12.1**Comparison between Traditional Calibration and Validation Strategy for Laboratory-Based Instruments and Process Sensors and Analyzers**

Experimental criteria	Laboratory-based instruments	Process sensors and analyzers
Number of measurements	Singlet or an average determination	Dependent on desired quality attribute of in-process material
Useful measurement	A measurement that meets a predefined acceptance criteria	A measurement that does not deviate not more than 3σ from the median of all measurements taken over a specified time period
Sample	Finished dosage form	In-process API, excipients or formulation
Standard	Instrument calibration standard or method reference standard	In-process material, in-process gas, liquid or inert substrate
Parameters	Invariant settings Instrument specific Depends on what method requires	Variable settings Process specific Depends on how process material varies over time
Performance verification	Performed prior to GMP measurement	Determine sensor/analyzer efficacy for in-process material Demonstrate sensor/analyzer signal and impulse ability and reliability
Efficiency	N/A	The ratio of the number of useful measurement results delivered by a sensor system to the total of all measurement results delivered over a specified time period
Design qualification (DQ)	21 CFR Part 11	QbD
Installation qualification (IQ)	cGMP	Development — SQ ^a Information only — limited Regulatory-filled release — SQ and vendor audit supplement
Operational qualification (OQ)	cGMP	Development — SQ ^a Information only — limited Regulatory-filled release — SQ and vendor audit supplement
Performance qualification (PQ)	cGMP	Development — SQ ^a Information only — limited Regulatory-filled release — SQ and vendor audit supplement

^a System qualification [17].**RTQA:** Real time quality assurance.**Sensor and analyzer ability:** The demonstration that the accuracy, precision, range and suitability of the sensor system or analyzer can be assessed from commission to retirement and yield consistent and predictable measurements not exceeding 3σ or less in any given specified time period.**Sensor or analyzer efficiency:** The ratio of the number of useful measurement results delivered by a sensor system to the total of all measurement results delivered over a specified time period.

SCADA: Supervisory control and data acquisition.

SPC Chart: The control chart is a technique for detecting which type of variation (“controlled variation,” “uncontrolled variation”) is displayed by a given process. The objective of the control chart is to give the user a guide for taking appropriate action to look for assignable causes when the data display uncontrolled variation, and to avoid looking for assignable causes when the data display controlled variation.

Transmitter: A device that sends a signal from a process location to a central location for control and monitoring.

Useful measurement: A useful measurement result is defined by a single measurement that does not deviate not more than 3σ from the median of all measurements taken over a specified time period.

REFERENCES

1. Ambrozic, C., and Parsons, J., *Real Time Quality Control (Batch monitoring) of Pharmaceutical Production Processes*, Umetrics Inc., Kinnelon, NJ 07405 (Presented at FDA, November 28, 2001).
2. Expanding the Vision of Sensor Materials, Committee on New Sensor Technologies: Materials and Applications, National Materials Advisory Board (NMAB) (1995).
3. Control Engineering Europe, June 1 (2001).
4. *Pharmaceutical cGMPs for the 21st Century: A Risk-Based Approach*, www.fda.gov/oc/guidance/gmp.html (2002).
5. *Current good manufacturing practice in manufacturing, processing, packaging, or holding of drugs: general* 21 C.F.R. Part 210 (2003).
6. *Current good manufacturing practice for finished pharmaceuticals* 21 C.F.R. Part 211 (2003).
7. *Guidance for Industry, PAT — A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance*, <http://www.fda.gov/cder/guidance/5815dft.htm> (2003).
8. Cohen, N., *Pharmaceutical Engineering*, Validation of a Water System for the 21st Century, July/August 1999, 1–6.
9. Ritchie, G. E. and Hashemi, S., *American Pharmaceutical Review*, GMP Roll Out and Implementation of Near Infrared Spectrophotometry for Pharmaceutical Testing, Spring (2003).
10. Cool, T., Proper definitions for Uncertainty and Risk, <http://www.dataweb.nl/~cool/Risk/ProperRisk.pdf>, July 25 (2001).
11. Shewhart, W. A., *Economic Control of Quality of Manufactured Product*, New York: D. Van Nostrand (1931).
12. Neave, H. and Wheeler, D., *Ninth Annual Conference of the British Deming Association*, Shewhart’s Charts and the Probability Approach, © 1996.
13. <http://www.cpac.washington.edu/NeSSI/NeSSI.htm>.
14. Barney, M., Motorola Inc., Motorola’s Second Generation, Six-Sigma Forum Magazine, May 2002.
15. Papadoyannis, I. N. and Samanidou, V. F., *Validation of HPLC Instrumentation*, Encyclopedia of Chromatography, June 2003.
16. Burns, D. A. and Ciurczak, E. W. (eds.), *Handbook of Near-Infrared Analysis*, 2nd edn, Revised and Expanded, Chapter 28 (2001).
17. Schadt, R., Process Analytical Technology — Changing the Validation Paradigm, *American Pharmaceutical Review*, 2004, 7: 58–61.

Part III

Methods Development

13 Sampling, Sample Preparation, and Sample Selection

Phil Williams

CONTENTS

13.1	Introduction	268
13.2	Sampling	268
13.3	Sampling and Error	272
13.3.1	The Source of the Sample	272
13.3.2	Sampling Itself	272
13.3.2.1	Sampling Location	272
13.3.2.2	Collecting the Sample	274
13.3.2.3	Grains and Similar Materials	274
13.3.2.4	Forages, Fibers, and Related Materials	277
13.3.2.5	Liquids	279
13.3.2.6	Semisolids	280
13.3.2.7	Nonsampling	280
13.3.2.8	Dry or Low Moisture Crystalline Solids	280
13.3.2.9	Sampling of Very Large Objects	280
13.3.3	Sample Handling	281
13.3.4	Blending, Subsampling, and Sample Identification	281
13.4	Sample Preparation	282
13.5	Grinders and Grinding	283
13.5.1	Types of Grinders	283
13.5.2	Factors Affecting Grinder Performance	284
13.6	Blending of Samples	287
13.7	Sample Storage	287
13.8	Sample Cells and Sample Presentation	289
13.8.1	Cell Cleanup	290
13.8.2	Static Electricity	291
13.8.3	Undetected Moisture	291
13.9	Sample Selection	292
13.9.1	Spectral Sample Selection	293
13.10	Summary	294
	References	294

13.1 INTRODUCTION

Near-infrared (NIR) technology offers a rapid method for the analysis of a wide range of material for a large number of chemical and physicochemical parameters. Both NIR reflectance and transmittance (NIT) have become recognized as rapid, accurate, precise, and pollution-free methods of analysis, and have earned respect among instrumental methods. The main reasons for the rapid climb to fame of NIR technology are (a) its speed while maintaining accuracy, (b) its freedom from chemicals and therefore from pollution, and (c) simple sample preparation. Modern NIR instruments are durable, and are capable of excellent accuracy and precision. They are backed up with comprehensive software and generally reflect a great deal of inspired input from the instrument company engineers. The main sources of error in NIR testing lie with the samples, their selection, preparation, and reference analysis. This chapter is an informal attempt at drawing attention to the chief trouble areas in these aspects of NIR technology. Provided they are efficiently accomplished, excellent analysis can be achieved with several different software systems. If they are not efficiently carried out, consistent and accurate NIR analysis is not possible, no matter how sophisticated the software. More recently, improvements in whole-grain instruments have revolutionized NIT and NIR testing of grains and seeds. This is done mainly to the removal of the need for grinding samples, and the errors thereby induced. Sample preparation, however, (including grinding) is essential to reference analysis, and many ground-sample NIR instruments are still in use.

But why do we need analysis at all? The objective of any chemical or physicochemical analysis is to determine the potential functional properties of the material being analyzed. Chemists and technologists in many fields have striven to develop tests that will enable them to predict end-product utilization potential of a substance from a knowledge of its chemical composition. The best known example is probably the protein test for wheat, which assists the cereal chemist in estimating bread and cookie baking quality. Moisture tests are carried out to determine whether a shipment of grain is safe to store or whether it should be dried, and also to determine for how much water is the buyer paying the full grain price. Nutritionists need information on energy, digestibility, and other parameters. But during the 30 or so years that have passed since the first introduction, NIR has transcended the confines of grain technology, has expanded into most fields of industry and commerce, and has made its presence felt in the areas of medical diagnosis and environmental science.

To determine functionality and price, primary and secondary processors and commodity brokers demand the most accurate chemical and instrumental analyses obtainable. But what of the material tested? This is almost invariably a sample of the complex conglomeration of flour, grain, cotton/polyester, cottonseed, silage, milk, coal, or whatever commodity is to be purchased or used. The integrity of laboratory analysis, whether it be by NIR or a reference *wet chemistry* method lies at the mercy of the sample tested. The “sample tested” is in turn dependent on the sampling method and its preparation for analysis. This chapter will address these aspects. Sample selection is also very important. In particular, it affects calibration of NIR instruments and is discussed later.

13.2 SAMPLING

About 30 factors affecting the accuracy and precision of NIR analysis are attributable to sampling, samples, or sample presentation. These are summarized in Table 13.1.

Most analysis is carried out on what is purported to be a sample that is completely representative of the whole. This can be achieved only by (a) continuous sampling of the shipment as it was delivered followed by accurate subdividing into a manageable amount for testing, or (b) repeated random sampling of the material followed again by accurate subdividing.

But what does our representative sample tell us? Accurate analysis of the sample will tell us the average or mean result of the test applied to the final sample. It will tell us whether the sample would meet specifications laid down prior to purchase. It will enable us to predict the likely functionality of the entire parcel of material represented by the sample in the process for which it is intended.

TABLE 13.1
Sample-Associated Factors in NIR Analysis

Sampling	Samples	Sample preparation
a. Type of sampler	a. Type of material	a. Type/Model of instrument
b. Location of sampler	b. Composition	b. Type of test
c. Material to be sampled	Oil	Intermittent
d. Foreign material	Moisture	In-/online
e. Physical nature of material	Fiber	c. Sample cell type
f. Size of sample	c. Physical texture	d. Sample cell size
g. Flow characteristics	d. Foreign material	e. Particle size
h. Sample transfer method	e. Blending	f. Bulk density
i. Blending	f. Identification/documentation	g. Composition
j. Storage		Oil
k. Identification/documentation		Moisture
l. Variability of population		Fiber
m. Frequency of sampling		h. Physical nature
n. Subsampling		i. Stratification
o. Sample selection (calibration)		j. Static electricity
		k. Cell loading
		l. Cell cleanup
		m. Grinder type

It will tell us the value of the sample, and therefore the price. But will it tell us everything we need to know? Not necessarily. The “truly representative sample” can be dangerously misleading.

A typical example of the unreliability of the truly representative sample concerns moisture testing. A continuous sample taken during the filling of a large bin at a subterminal elevator, wherein the grain is likely to remain in storage for several months, may indicate that the average moisture content of the grain in the bin is 12.8%, low enough for safe storage. In practice, the moisture content of subsamples may range from less than 10% to over 16%. The pockets of higher moisture grain would be likely to support the development of fungal infestations so that when the grain is removed from storage, an extensive degree of mustiness and spoilage may have occurred, and the functionality and value of the grain could have fallen significantly. As well, heat may develop in patches where the moisture content was high, to the extent of damaging the grain.

When a composite sample is taken to represent the entirety, the relevant tests should be carried out on all subsamples that contribute to the final representative sample in order to find out what the sample truly represents. It is important to know the variance within a population, such as a grain bin or stack of bales or a batch of cans, as well as the mean analytical results, so the range and standard deviation should be reported as well as the mean. The mean itself should be reported as a weighted mean based on the weight of material as well as on the composition represented by each subsample. This is particularly true of NIR testing, because the reliability of calibration lies in the natural variance incorporated into the calibration samples. On the other hand, the simplicity and speed of a NIR test of the subsamples facilitates testing all subsamples of a sampling sequence. Using NIR technology, it is possible to report data for each individual subsample of a composite. NIR/NIT instruments can greatly improve the integrity of sampling and analysis. If NIR analysis of every subsample shows a high degree of variability in moisture content, the bin contents should be turned if the grain is intended for long storage, to redistribute the moisture and thereby reduce the moisture content of the high-moisture patches.

A note here about *weighted means*. If subsamples are to be analyzed and compiled into a report on the overall composition of the entire population, it is sometimes necessary to relate each analytical

result to the actual amount of material represented by the subsample. For example, a cargo of wheat may be subsampled eight times during the loading. The samples may represent different amounts of wheat, depending on things such as the time of day, etc. The overall result should be compiled into a weighted mean. An example is given:

Report on Protein Content of the Marine Vessel “Ocean Maru”*

Subsample	Weight in tons**	Protein content	Weight × protein content
1	4,326	13.9	$4,326 \times 13.9 = 60,131.4$
2	5,651	13.2	$5,651 \times 13.2 = 74,593.2$
3	4,894	13.5	$4,894 \times 13.5 = 66,069.0$
4	5,904	13.6	$5,904 \times 13.2 = 77,932.8$
5	5,012	13.7	$5,012 \times 13.7 = 68,664.4$
6	4,655	13.4	$4,655 \times 13.4 = 62,377.0$
7	4,310	13.8	$4,310 \times 13.8 = 59,478.0$
8	1,255	14.5	$1,255 \times 14.5 = 18,197.5$
Total weight	36,007		
Mean protein %		13.68	$(13.9 + 13.2 + \cdots + 14.3)/8$
Weighted mean protein %		13.54	$(60,131.4 + 74,593.2 + \cdots + 18,197.5)/36,007$

* Fictitious name.

** Weight represented by subsample.

In this case the entire cargo was guaranteed to contain wheat at 13.5%. The weighted mean showed that after loading the first 30,000 tons the mean protein content was 13.47%, in other words below the guarantee. The addition of 4310 tons at a higher protein content brought up the level to exactly 13.50%. The cargo was then “topped up” or “finished” with high-protein wheat to ensure that the guarantee was met.

The total yield of a single plant may be regarded as a truly representative sample, because there is no more grain available. Nondestructive tests such as kernel weight or NIT tests for protein or moisture will tell the tester the average result for that plant, but again individual seeds will vary in size and composition, and the variance is still an important factor in the individuality of the plant. For example, the seeds in pods of one faba bean genotype may differ from top to bottom of the plant to a much greater degree than those of another genotype.

In the field of NIR technology, there has been a tendency to think of samples in terms of grain because grain is what originally focused attention onto NIR as a rapid and accurate method of analysis, comparable to other physicochemical methods [1]. In the early days of NIR testing, more than 95% of all applications dealt with grain or products immediately derived from grain, such as flour, semolina, or malt. The last half of the “NIR era” (1985–2000) has seen its rapid diversification into many fields in food, agriculture, plastics, pharmaceuticals, petrochemicals, textiles, wine, dairy products, forages, mixed feeds, and other areas, including live animals and people, and the terrain (lakes, rivers, and seas) that makes up the environment.

A sample can be anything from a few grams of grain or a few milliliters of beer to a fistful of straw, a sheet of plywood, or a human arm. Liquids with viscosities ranging from thinner than water to thick slurries; plastic solids such as plastics themselves or something of the consistency of cheese; a piece of cloth, soil, wood, or wool; an area of skin; a liter of river water; or a bagful of cow manure, all of them present individual precepts for either bringing the material into contact with the NIR instrument or, more recently, the NIR instrument to the sample. All of these materials

pose the questions of (a) how to take the sample, (b) how to prepare it in a manner that is acceptable to the instrument and that will assure reliable results, and (c) how to present it to the instrument.

Sampling methods can be subdivided into three types: (a) automatic, (b) manual, and (c) no sampling at all as such due to (i) the entire population being tested or (ii) the NIR instrument reading the sample directly as in online systems or by interactance [2]. An important aspect of sampling for NIR work is that practically any application calls for calibration of the instrument against a reference method. Samples for NIR analysis should be withdrawn and prepared in exactly the same way as those used for the reference analyses used for calibration. Exceptions lie in some industrial applications and in interactance applications. When the material to be analyzed is too big or too inaccessible to furnish 60 to 100 samples for calibration and verification, sample sets may be artificially created by varying composition of a suitable number of small batches of the material. These can be used to calibrate NIR equipment, and the accuracy of the subsequent analysis verified by determination of the functionality of the end product. Often in such applications, the constituent to be determined is either a “standout” type of chemical compound or water, and there are generally fewer interferences than are encountered with agricultural and food applications.

For most applications, it is essential that the state of all samples be identical when the results of the reference analysis are used in calibration. This means that physical parameters such as particle size and shape of the calibration test, and reference samples should be mutually compatible, and that the reference results must be relayed to the NIR instrument on the same “as-is” moisture, oil, or more simply constitutional basis that the instrument will see. The formula for conversion of the percentage of a constituent to an as-is moisture basis from data reported on a constant moisture basis is:

$$\begin{aligned} \% \text{ composition as is} &= \% \text{ composition} \times [(100 - \% \text{ moisture as is}) \\ &\div (100 - \% \text{ moisture constant})] \end{aligned}$$

For example, a series of samples have been submitted for calibration with protein content (Dumas) values reported on a 12% moisture basis. One sample has a moisture content of 9.1% as is and a protein content (12% moisture basis) of 13.7%. What is the as-is protein content?

$$\begin{aligned} \% \text{ protein as is} &= 13.70 \times [(100 - 9.1)/(100 - 12)] \\ &= 13.70 \times (90.9)/(88) \\ &= 14.15\% \end{aligned}$$

This may be complicated in the analysis, for example, of fresh meat, which may contain up to 80% water, due to the difficulty in determining the water content. An error of 2% in determination of the moisture can lead to errors of over 1% in reporting the protein and fat contents. The same criteria apply to fresh fruits and vegetables, and to materials such as silage or wood pulp, fresh breads — in fact, to any high-moisture material. This is particularly important when the reference analysis depends on 1 g or less of finely divided material. NIR test cells may hold several grams, while whole-grain instruments may test 100 g or more. Moisture can cause lots of problems in calibration and verification — it may even be the biggest single source of trouble — but we aren’t going to talk about calibration here!

Errors in NIR analysis are frequently credited to the reference analysis method, whereas the transgression has often occurred before the reference method is begun. In a Kjeldahl test, the assumption is made that sample preparation stops before the laboratory; after all, the weigher weighs a prescribed quantity of the sample sent to the laboratory. But the test is really determining total nitrogen in the form of ammonia. The actual measurement is done at the titration bench and only takes a few seconds. Everything up to there, including digestion and distillation, is sample preparation. The actual determination of total nitrogen by titration is just as fast as an NIR test, but the sample preparation takes much longer. This aspect is particularly important in some applications where it is both easier and more practicable to use NIR than a standard method.

Sometimes NIR instruments can make a measurement of a parameter that is more meaningful than a reference chemical or physicochemical test. An example is the testing of wheat for hardness. Hardness affects the way in which wheat breaks down during grinding, and hard wheats have a higher mean particle size (MPS) and more coarse particles than softer wheats. Classical hardness tests depend on sieving or sedimentation. Sieving is essentially a two-dimensional measurement, while sedimentation methods mainly depend on a derivation of Stokes' law, which assumes all particles to be spherical. NIR is a three-dimensional measurement that takes cognizance of three-dimensional size and shape, and is therefore theoretically a more comprehensive principle for assigning a hardness index to wheat.

13.3 SAMPLING AND ERROR

There are four main sources of error associated with the sample. These are (a) the source of the sample, (b) the sampling method, (c) the sample itself, including sample preparation, and (d) final sample presentation.

13.3.1 THE SOURCE OF THE SAMPLE

What does this mean? One concept is that the sample has to represent the whole population, such as a cargo of grain, a collection of bales of wool, or the soil of a field. Another concept is that the sample should provide information on what the operator needs to know. This may be furnished by taking samples of materials that are, or have been affected by certain circumstances. This type of sample would include samples drawn from a material during processing. The analysis would provide information as to the efficiency of the process, rather than relating to the actual value of the material being sampled. Another example is the analysis of plant or animal material, to provide information on fertility, or the impact of a change in the environment. Animals such as mussels live for several years at the bottom of lakes and rivers, and respond to changes in their environment, such as those induced by construction of dams, etc. Analysis of these over a period can yield valuable information on the progression of deterioration or recovery of their growing location.

Soil analysis is beset with the further complication that samples have to be taken from different depths, to reflect movement of nutrients during the seasons. *Precision agriculture* is focusing renewed attention to the selection of methods of sampling fields, in order to improve the efficiency of fertilizer use. Here the controversy is whether analysis of plant material is more practicable than soil analysis in provision of the necessary information. Plant material, such as grain or herbage, is more readily applicable than soil analysis, as well as being more amenable to analysis by NIR. It also reflects nutrients that are available to the plant.

13.3.2 SAMPLING ITSELF

About 14 factors affect the actual sampling process. These include (a) identification of the area or areas in the operation from where the sample or samples are to be collected, (b) collecting or taking of the sample, (c) the nature of the material to be sampled, (d) the amount and type of foreign material present, (e) the physical size, shape, and bulk density of the individual items comprising the population, (f) flow characteristics, (g) size of sample required, (h) packaging the sample, (i) transporting the sample, (j) blending the sample, (k) sample identification, (l) natural variance within the population that will affect the frequency of subsampling, (m) subsampling, and (n) storage of the sample.

13.3.2.1 Sampling Location

Deciding on the most suitable location at which sampling is to be carried out is not always a simple process. In a large grain terminal, it is customary to take the sample during horizontal rather than



FIGURE 13.1 Hand sampling a farm truck at a country elevator.

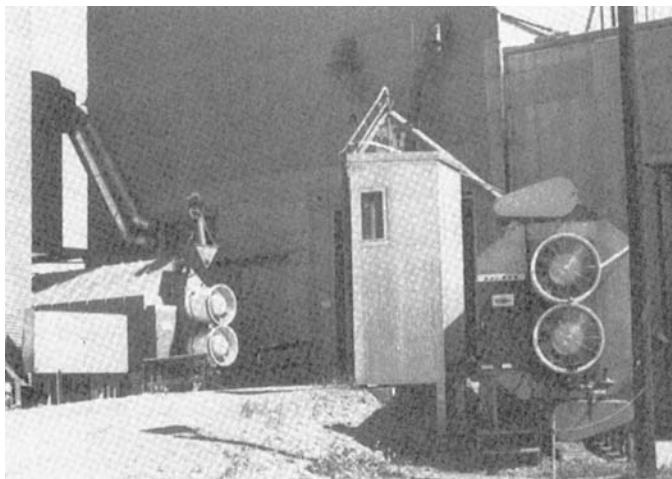


FIGURE 13.2 Pneumatic probe for sampling farm trucks.

vertical conveying. Belt and chain conveyors lend themselves more readily to installation of sampling devices. In a primary (country) elevator, the sample may be taken manually at the back of the truck during unload (Figure 13.1) or a pneumatic probe may be installed (Figure 13.2).

The location of the probe, access of the trucks to the unloading ramp, and transport of the sample have to be taken into consideration. In loading railcars from primary delivery points, the grain often drops through the elevator en route to the scale and the leg that elevates the grain into the discharge spout. The vertical stream of grain can be sampled manually or automatically. Sampling of railcars or trucks during loading from a primary (country) elevator can be carried out by manually scooping a sample periodically from the stream of grain that passes down through the elevator before entering the car. This system, and sampling from the rear of a farm truck during unload into an auger are examples of sampling vertical streams. Sampling fields for standing crops or soil involves a comprehensive identification of sites within the field to ensure statistically correct coverage. Soil sampling is further complicated by the need to sample at two or more depths. In industrial plants, location of sampling has to embrace both raw materials and finished products, and

often intermediate products. Temperature and physical state of the sample, accessibility of sampling and transporting equipment, and environmental conditions at possible sampling locations are factors affecting the decision as to where to locate the sampling system. It is often necessary to locate sampling at several stages of production to determine progress in the development of the product so that changes in composition of product and raw materials or conditions may be monitored or implemented. Sampling for environmental work involves sampling waters of lakes, rivers, streams, and oceans, also the sediments underlying these bodies of water. Special equipment may be required, and even specialized staff, such as divers, to take samples of sediments and living organisms.

13.3.2.2 Collecting the Sample

This is related to [3] the type of material to be sampled.

13.3.2.3 Grains and Similar Materials

Foreign material, insects, etc. are usually present in grains and seeds, and form part of the grading/pricing system. It is essential that any sampling system results in samples that represent the amounts of these elements present accurately. Systems that bias the amounts of light or heavy foreign material, or kill insects, etc. are not acceptable. Automatic sampling is widely used in the grain and food industries, and because of its continuous operation provides excellent service to the testing laboratory. There are several forms of automatic sampling. The diverter-type sampler used in modern terminal grain elevators provides a comprehensive sample of a truckload, a railway carload, or a ship, both loading and unloading, and can also be used to provide a running sample of a bin full of grain during the filling or emptying of the bin. Grain moves horizontally into and out of elevators and warehouses on conveyors, either belt- or chain-type, either of which can be sampled with a diverter sampler. The principle of the diverter sampler is illustrated in Figure 13.3.

The sampling arms pass through the grain mass periodically, and the grain is delivered pneumatically directly into the inspection or testing laboratory. The amount of sample taken is regulated by adjusting the frequency by which the sampling arms pass through the grain. An essential adjunct to the diverter sampler is the stream splitter or sample divider (Figure 13.4). Diverter samplers usually deliver a relatively large amount of sample, too much for testing purposes, so that some of the sample must be redirected back to the conveyor.

The stream splitter must be very accurate, and the sampler arms must dip to the bottom and reach the sides of the conveyor in order to ensure even distribution of grain, including foreign material. Flour and other powdered substances can also be sampled using specialized diverter samplers. These work using an impeller that receives the free-flowing materials from conveyor tubing, accesses it to the viewing area of the NIR equipment, and effects its ejection and return to the main stream, if necessary, or to a receiver for laboratory verification (Figure 13.5). Size of the sample is regulated by the size of the impeller, and frequency of sampling by a timing system.

Another type of automatic sampler is the Woodside sampler, which consists of chains carrying small cups (Figure 13.6). The chains pass through the grain stream, the buckets fill and empty directly into a hopper, from which the sample can be transported by gravity pneumatically or manually into the test laboratory. The size of sample delivered by a Woodside sampler can be regulated by the number and size of the cups. The modern tendency is to move toward the diverter-type sampler. These are generally more flexible than the Woodside sampler. For example, peas are very hard and more or less round, and tend to bounce into and out of the small cups of a Woodside sampler. Diverter samples are not subject to this source of error.

It is practicable to withdraw too much sample, then subdivide it using a stream splitter. The subdivision step helps to assure that the sample is well blended before it is analyzed. Pneumatic transfer over long distances may cause changes in moisture content due to the amount of air used, and the constant abrasion of sample against itself and the walls of the conveyor tubing. But NIR

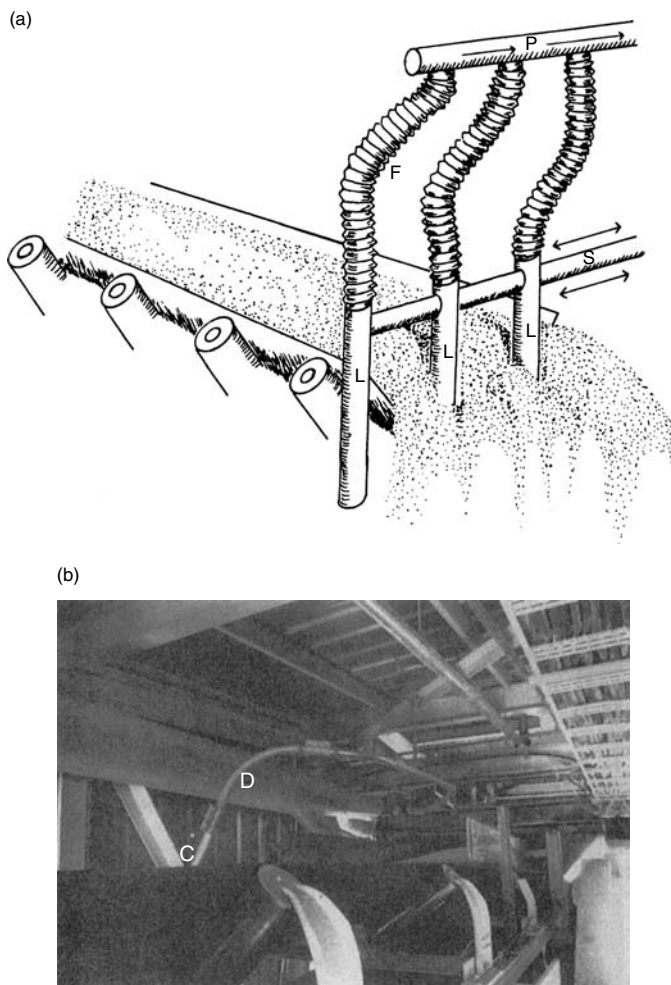


FIGURE 13.3 Diverter sampler. (a) Principles. Probes (L) move through grain stream by the action of reciprocating arms (S). Grain is moved through flexible tubing (F) pneumatically to receiving tubing (P) and thence to the inspection/analytical laboratory. (b) Diverter sampler *in situ*. C = sampling probes. D = conveying tube to laboratory. (Courtesy Pioneer Grain Ltd., Vancouver, Canada.)

technology protects the integrity of the analysis by affording the means of testing for moisture simultaneously with everything else and enabling analytical data to be reported on a constant-moisture basis. Automatic samplers can be used for any free-flowing material, including liquids, where the sampling devices can be simply operated by suction. The sampling method is governed principally by the nature and bulk of the material to be sampled.

Many variations on the previously mentioned themes have been developed for use in the grain, food, dairy, animal feed, fermentation, petrochemical, and other industries, and the efficiency of all of them depends on the five premises of withdrawal, blending, subdivision (optional and not always necessary depending on the installation), transfer, and preservation of the initial condition of the material. The efficiency of the system in preservation of the sample during the operation should be carefully checked by manual sampling and transfer prior to full-scale adoption.

There are many instances where analysis of, say, a cargo of grain after delivery for protein content shows a consistent 0.1 to 0.3% difference from the mean of the samples taken continuously during loading. This may be due to a consistent difference in moisture content between grain delivered

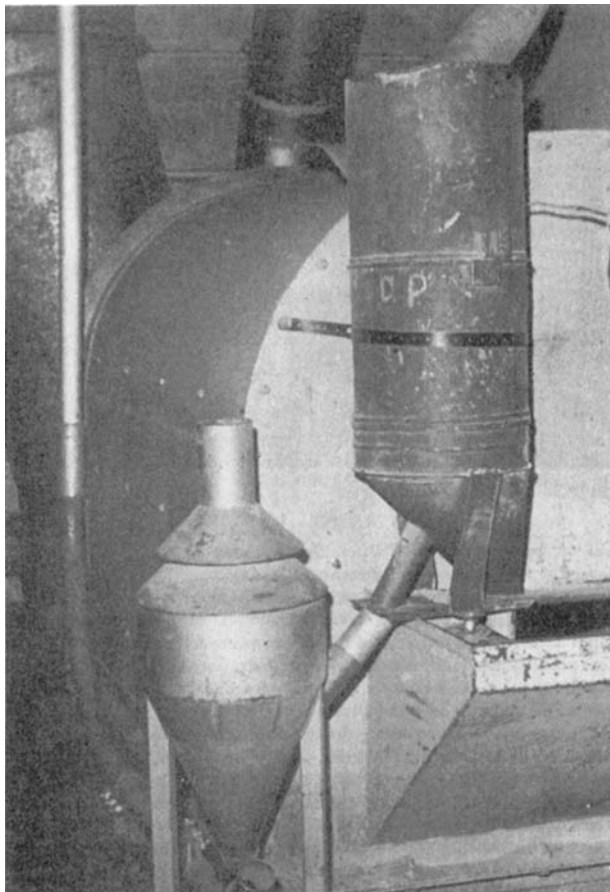


FIGURE 13.4 Sample stream splitter. (Courtesy Saskatchewan Wheat Pool Ltd., Vancouver, Canada.)

by bucket elevator and belt to a ship, and that of the grain sample transferred over a fairly long distance pneumatically to a testing laboratory. The pneumatic conveying may cause a loss of 1% moisture or more, relative to the more gentle conveying of the main grain stream. Differences in infestation, or foreign material content can also occur due to imperfect sampling. Adjustments can be made to fine-tune sample size and ensure that practically no changes take place during automatic sampling. The checking should be repeated from time to time, particularly if operational conditions vary significantly due to weather, workload, and other factors. Bins at grain terminals can be sampled at any level by vacuum-operated probes, which can penetrate to depths of up to several meters, and are useful for sampling bins where grain has been held static for a long period. The samples can be subdivided after withdrawal, by equipment such as the Boerner divider.

Manual sampling is gentler than automatic sampling. It is also much slower, and subject to the error of complacency. In some situations such as hand sampling from moving conveyors, or streams of grain falling from appreciable heights by gravity into bins or holds, hand sampling can be hazardous due to the possibility of injury or dust inhalation. In the grain industry, hand sampling is carried out mainly on delivery of farm trucks to an elevator or processing plant and during the loading of railcars or trucks from a primary elevator. The sampling may consist simply of diverting grain into a pail by hand. Provided the sampling is carried out during the entire unloading or loading operation, an excellent sample can be taken.

In some developing countries that import grains, hand sampling of belts is fairly widely encountered. The position available to the hand sampler may result in only part of the conveyor

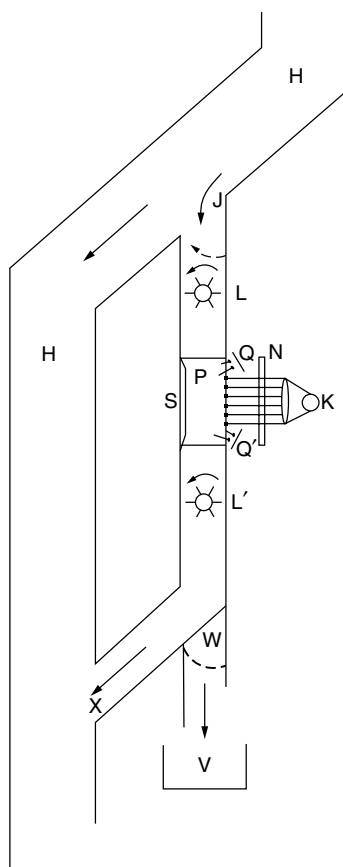


FIGURE 13.5 Flour/grain impeller continuous sampler for NIR analysis. H = main flour stream; J = access for flour samples (covered by electronically controlled flap); L, L' = impellers; K = light source; N = collimating lens; P = sample sensing cell; Q, Q' = detectors; S = compression system; W = sample egress, controlled electronically; V = flour sample receptacle (for laboratory verification); X = flour return to main stream.

being accessible, and samples can be very unreliable (a) if the sampler has not been thoroughly trained in the principles and technique of sampling, and (b) if there is not clear access to all parts of the conveyor. Hand samplers may take the form of hollow probes of various lengths for probing bins (Figure 13.7), truckloads, or railcar loads; sampling cups such as the Ellis cup for sampling belt conveyors [3] and various types of sampling cups with longer handles such as the Pelican sampler used for sampling streams of grain.

Another form of hand sampler is the bag probe, used for probing bagged commodities. The bag probe is a small metal trough open at one end with a point at the other. The point is thrust into a bag and grain flows out down the trough and is collected in a suitable receptacle. Subsamples obtained from probing several bags can be blended to form a composite sample, but for the most complete information each individual bag should be tested at the upper and lower halves of the bag.

13.3.2.4 Forages, Fibers, and Related Materials

Bales, stacks and pits of hay, other forages and silages can be sampled by metal probes that consist of steel bars of various lengths, the ends of which carry several barbs. The probe is pushed into the bale or stack, twisted, and withdrawn. The barbs carry a sample from the inside. Several bales should be

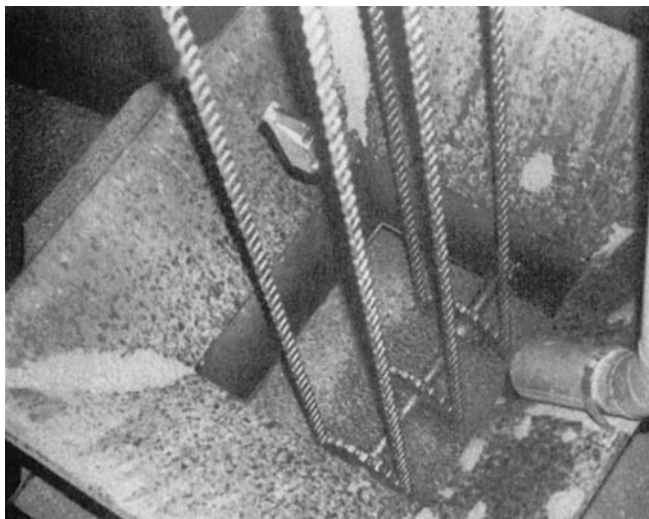


FIGURE 13.6 Woodside sampler.



FIGURE 13.7 Hand probe for sampling railcars or trucks.

probed or several probes made of a stack or bin at different levels and locations to make up the final sample. Again it is advisable to analyze as many subsamples as possible, to determine the variance that exists within the bulk of material.

Bales of wool and cotton may also be sampled by a probe or core type of sampler. The Wool Research Organization of New Zealand developed a very efficient automatic core-type sampler for very dense wool bales. It also perfected a subsampler to provide samples for direct NIR analysis for residual grease and moisture. Woven textiles can be sampled by cutting out a piece of the fabric or, more simply, by taking a direct reading of the fabric. Care should be taken with direct NIR reading of woven and other fibrous textiles, because the direction and orientation of the weave may affect the diffuse reflectance and accuracy of analysis. Sampling of paper is similar to that of textiles. Breads can be sampled by slicing and either analyzed directly or after drying and grinding. Fruits and vegetables are sampled by taking whole fruits or vegetables. Sample presentation involves subsampling individual items of the fruit or vegetable.

A word about reporting data for nutritional purposes. When materials with high moisture content, such as fresh fruits and vegetables, meat, fish, or cooked items, such as canned peas, or even breads, are to be used in formulation of diets or feeds, the composition may have to be reported on an as-is basis with respect to moisture content. For example, the protein content of peas is well known to be about 27 to 30% on a moisture-free basis. But nothing or no one eats dry peas — they are much too hard, and they are cooked before eating. The moisture content of cooked peas is about 67%, so the protein content of the peas as they are actually eaten is reduced to about 9 to 10%. This is the contribution that the peas make to the diet, where the diet or animal feed is to be prepared to contribute a certain amount of protein, etc. per day.

13.3.2.5 Liquids

Sampling of liquids depends on the viscosity and clarity of the liquid. Very free-flowing liquids such as gasoline, wine, beer, worts, liquors, whey, and others can best be sampled and analyzed simultaneously using flow-through cells. This system can also be applied to solutions of substances, such as extracts. The principles of transmittance [4] or direct transmittance are most widely used. Milk presents a special condition, because it consists of a suspension of fat globules in an aqueous solution containing protein, sugar, and other substances. The fat globules are lighter than water, and are in more or less constant movement. This constitutes an important source of error in NIR transmittance spectroscopy (NIT), because the fat globules cause variable diffusion of the light as it passes through the sample.

Transmittance is not likely to be more successful than transmittance for analysis of milk by NIR, because the surface presented to the instrument will be changing in composition, due to the progressive changes in fat content. Transmittance or reflectance of the sample in a large cell, with continuous agitation is the recommended method of sample presentation for NIR analysis of milk, or any other type of liquid sample containing suspended material that is liable to settling at the bottom or surface of the cell.

As viscosity increases it becomes increasingly difficult to use flow-through cells, due to the pressures necessary to force the heavier liquids through the fairly fine tubing used in flow-through cells, and also to difficulties of sample-to-sample clean-up. Thicker liquids such as lubricating oils or cosmetics are best sampled by withdrawing the sample with a scoop or syringe and analyzing it in a specially designed cell. As liquids become thicker they approach the slurry consistency. Temperature exerts an important effect on viscosity, and it is important to analyze the sample at the same viscosity as when the sample was withdrawn. Differences in viscosity may affect refractive index and, in turn, the diffuse reflectance from or transmittance through the sample.

The next degree of viscosity includes extremely viscous materials such as face creams, some salad dressings, and epoxy and other types of resins. These can be sampled in the same way as lubricating oils. Opaque types of slurry are best analyzed by diffuse reflectance. Some industrial

slurries are at elevated temperatures when sampled but solidify on cooling. This type of slurry can be poured into a cell, a surface “struck” while the sample is cooling, and the smooth surface of the cooled solid used in NIR analysis. This type of sampling is used in the cocoa industry. Cocoa butter and chocolate itself can be sampled from melts in this way.

13.3.2.6 Semisolids

Ordinary butter (the toast type) is another example of a liquid/semisolid, depending on temperature. Butter is really a slurry of water in fat, differing from many materials of similar stiff room temperature textures in that its production is essentially a room temperature operation. Materials of this type can be sampled with a core-type sampler in the semisolid state or by methods previously described, in the liquid state. They can also be sampled by an extrusion system, whereby the sample is “chopped off” from an extruded stream. It is important that temperature be controlled so that the cut surface remains level and does not start to flow. Cheese is even more solid and can be sampled by simply taking a slice. “Processed” cheese is usually manufactured in the liquid state, and can be sampled and analyzed as-is, or fed into a sampling cell and allowed to solidify before analysis. “Crumbly” materials, such as biscuit (cookie) or cake doughs, dough mixes for pasta or noodle production or compressed (fresh) yeast, can be sampled manually or by an extrusion system. This includes thick slurries, which will eventually either solidify or be compressed into solid sheets or blocks.

13.3.2.7 Nonsampling

This method can be applied to NIR analysis of large sheets of material. This includes some types of plastics, paper, cardboards, and materials such as certain types of wall board. These materials can also be “sampled” by “nonsampling.” Nonsampling means that the NIR instrument reads or analyzes the sample by direct application of the detection system to the sample surface and communication of the signal to the computing system by means of fiber-optics. The signal is taken by diffuse reflectance from denser materials or by body transmission from less dense surfaces. Body transmission or interactance is a technique invented at the U.S.D.A. laboratory at Beltsville, MD, and recently introduced for the NIR determination of body fat in human beings [2]. The sensing unit is applied to the surface and the signal reaches the detector from around the sensing unit.

Materials such as wood, rubber, solid plastics, and other materials can also be nonsampled and analyzed directly. NIR analysis of this type is simplified by the fact that usually only one or two constituents are to be determined. Also there is a very strong signal at specific wavelengths, relative to the background, for the constituents to be determined. Conventional sampling of these materials must be used for reference analyses. Fiber-optics can now extend the range of offline measurements to several meters, and it is likely that both fiber-optics and interactance will find many applications in NIR analysis in industry in the future.

13.3.2.8 Dry or Low Moisture Crystalline Solids

Materials such as tablets of pharmaceuticals, coals, certain salts, soil, and similar substances, can be sampled manually or automatically. Soil is almost invariably sampled manually. At the other extreme, substances such as raw meat and fish, which have very high moisture contents, are also usually sampled manually. One reason for this is that the parent materials of substances such as meat or fish are individual animals.

13.3.2.9 Sampling of Very Large Objects

Objects, such as trees, or aquatic marine plants present the questions of what part of the plant should be sampled. This may require research to determine the areas of the object, the composition of

which is most closely related to the functionality (of an object such as a tree), or status (of a lake, river-bed, etc.). The question of sample preparation of materials such as water hyacinth also has to be addressed, because these plants are very large. Operations such as drying can only be carried out on subsamples, and it is important to determine the areas of the plants that will yield the information required. Large animals do not present such a problem, since organs, body tissues, or fluid can provide the information.

No matter what the material or sampling method, the variance between individual samples or subsamples of the same material should be established and recorded in order for the sampling technique to truly represent the composition and end-product utilization potential of the material being analyzed. This applies to situations as varied as grain entering a ship or a collection of mussels on a riverbed.

13.3.3 SAMPLE HANDLING

Sample handling includes packaging and transportation. Packaging of a sample is only a factor if the sample has to be transported an appreciable distance from source to laboratory, or if storage is necessary. In most cases, for grain, forages, and a wide range of commodities, plastic bags are very suitable for interim sample packaging. For liquids, plastic, metal, or glass containers may be used. If the sample is hot, some precaution should be taken to prevent implosion of the container during cooling. Many types of modern plastic containers come with tight-fitting lids, which effectively protect the contents from moisture interchange. These are inexpensive and available in square or round forms. Square containers are stackable, but round containers are easier to clean and are the most practicable. Care must be taken in stacking samples in square containers, so that the weight of containers stacked on top of each other doesn't cause breakage.

It is important to check the suitability of plastic containers from several aspects. For example, the lids can be too tight fitting, so that in opening the containers damage can be incurred to both the hands and the lids. (Damage to the lids is the more important, because they may not be reusable, whereas the hands will heal in a few days!) If samples are to be stored frozen, care should be taken in selection of plastic containers, because some plastics become brittle at freezing temperatures, and may crack during storage, or immediately on removal from storage. Shape and size of containers should also be evaluated for suitability.

Transporting the sample may be done automatically or manually. Manual sample transfer simply means carrying the sample to the laboratory or mailing it. Automatic sample transfer is usually done by means of pneumatics or by gravity. Both of these methods concern solid samples. Liquids and slurries can be pumped and, except for thick or "lumpy" slurries, can be analyzed directly by NIR using flow-through reflectance or transmittance cells. Pneumatic sample transfer is widely used for grains and other commodities. It is fast, and pneumatic ducting can be adapted to a wide variety of situations and distances. A disadvantage is that the use of high-volume air may result in some moisture loss. This should be monitored during the initial installation by cross-checking using hand sampling and sealed containers. Pneumatic transfer is fairly expensive in operation. Gravity sample transfer is also widely used. Automatic samplers can deliver into ducting that leads the sample directly into the laboratory. Hand-sampling systems can employ either type of sample transfer. Transport of samples to or from laboratories is often done using special envelopes or "Zip-lock" bags.

13.3.4 BLENDING, SUBSAMPLING, AND SAMPLE IDENTIFICATION

Before sample preparation the sample should be thoroughly blended. In the case of grains and powdered substances, liquids, and slurries, this reduces the possibility of stratification of the material, which can be a serious source of error. Stratification can occur in liquids due to differences in temperature, refractive index, composition, density, and other factors. Blending of forages means ensuring that the correct proportion of leaves to stems is maintained, and that the sample submitted

for sample preparation contains the correct proportion of plant species present in the sward. Blending of materials such as forages is best done by hand. Grain can be blended by hand, or by equipment such as the Boerner or Dean-Gamit sample dividers.

The efficiency of blending should be monitored during the development of a sample preparation process.

Subsampling is usually a necessary preliminary to sample preparation. This is usually carried out manually for “long” samples of forages, straws, and silages, and both blending and subsampling may be facilitated by using a small chaff cutter or similar chopping equipment to reduce the sample to fairly short pieces (1 to 2 cm in length). Care should be taken to minimize moisture, sap, and leaf loss. Granular commodities can be subsampled automatically with stream splitters, or manually with equipment such as a Boerner or Dean-Gamit divider, which can be used with a wide range of materials.

Sample identification is an important aspect of sampling. No matter how efficient the sampling method, all is in vain unless every sample is correctly identified. About 3% of all major errors in laboratory testing are attributable to analytical results being associated with the wrong sample. The durability of markers should be checked, to make sure that the identification doesn’t get rubbed off, or fades between sampling and analysis. The ideal marker is one that is stable to heat, time, and smudging, non-water-soluble, but removable with a solvent such as alcohol, to facilitate reuse of the containers.

13.4 SAMPLE PREPARATION

Sample preparation is the actual generation, synthesis, or preparation of samples for calibration and monitoring. It should be achieved in such a way that the sample presented for analysis has not been changed from its original state. Sample preparation without causing changes in moisture content may not be possible. Precautions should be taken to determine the moisture content before and after preparation, and the analytical data reported on the basis of the original moisture content, except in cases where data have to be reported on a constant moisture basis.

Sample preparation is the second fundamental principle involved in successful NIR analysis. It includes identification and documentation, cleaning, drying, subsampling, grinding, and blending. Some of these operations have already been described. Factors involved in sample preparation include the nature of the material including physical size, texture, etc., its composition, amount and type of foreign material present, grinding or other forms of size reduction, blending, and identification. Cleaning means removing foreign material that may interfere with analysis and with industrial processing. It includes removal of weed seeds, broken seeds, pieces of plant material, pods, stones, metal, soil, etc., from cereals, pulses, and oilseeds; removal of soil, etc., from forage plants; removal of vegetable matter from wool, and cotton; filtration of insolubles from liquids; and other forms of cleaning. A wide range of specialized equipment is available for cleaning grains and oilseeds [5,6]. Equipment for cleaning of other materials is specialized and often custom-made. The presence of large amounts of foreign material can affect the efficiency of withdrawal of samples. If it is not removed, foreign material may also affect the accuracy of analysis, because the composition and texture of the foreign material may be quite different from the sample.

Drying is necessary in many high-moisture materials such as fresh corn, silages, and other substances, before grinding and before analysis. Drying of grains and forages is extensively documented [7–9]. Care has to be taken to avoid damage to the physical nature of the material or to its chemical and physico-chemical makeup and functionality. Fresh corn is often harvested at well over 30% moisture and must be reduced to 15% or lower before grinding or storage. This can be achieved by forced air drying, ideally at temperatures of not higher than 40°C. Silages contain significant quantities of volatile fatty acids and esters that can be lost on drying, with concomitant changes in chemical makeup. Often the volatiles need to be determined, and losses on drying can cause serious errors in analysis. Losses of volatiles can be minimized by methods such as two-stage air drying, which includes a low-temperature (up to 35°C) stage A to a moisture level low enough to enable

grinding (12 to 15%), at which stage the volatiles can be determined. Moisture is determined in the partially dry material by conventional air oven drying at 130°C for 65 min (stage B). The volatiles are subtracted from the (moisture + volatiles) in the second stage, and the second-stage moisture used to compute the total moisture content using the formula:

$$\text{Moisture \%} = \text{Moisture A} + \left[\frac{(100 - A) \times B}{100} \right]$$

The volatiles must then be corrected to the moisture level at which the results are to be reported or to the original as-is moisture basis.

Alternatively, a known weight of the high-moisture material can be extracted with 95% ethanol, which will extract volatile fatty acids and most of the water. The water remaining can be determined by direct drying. The volatiles can be determined in the alcoholic solution by techniques such as gas-liquid chromatography (GLC) or high-performance liquid chromatography (HPLC). Another subsample of the original material can be oven dried and the total loss in weight corrected by subtracting the volatiles percent to give the total moisture content. Slurries can be analyzed directly, then moisture determined by a single- or two-stage method to enable accurate reporting of analytical data on an as-is or constant moisture basis. Volatiles can also be determined by a combination of oven drying and determination of moisture by Karl Fischer titration [10]. The volatiles are given by difference between oven and Karl Fischer results. The same applies to substances such as butter and cheese for which data on volatiles may be necessary.

Freeze-drying, or high-vacuum drying offers an alternative method of drying without loss of volatile substances. Freeze-drying is time-consuming, and expensive, but effective. Methods for drying material without loss of volatile constituents are painstaking procedures, but one of the objectives of this chapter is to provide advice as to the methods of sampling and sample preparation likely to give the most accurate analysis by NIR or standard methods. When the NIR instrument is calibrated to all constituents including moisture, think of all the trouble it will save!

13.5 GRINDERS AND GRINDING

In the early days of NIR analysis, sample preparation usually involved grinding a sample of grain into meal. Attention has been drawn to the influence of grinder type and grain type on the granularity of the resultant whole-meal [11]. But even a simple process such as grinding a sample of grain poses several questions. What type of grinder should be used? Is a certain type of grinder best suited to a particular commodity? What does the action of grinding do to a sample in terms of its composition? What are the main factors affecting the efficiency of a grinder? What are the criteria of a good grinder?

13.5.1 TYPES OF GRINDERS

There are five main types of grinders. The first is the burr mill, typified by the Falling Number models KT30 and KT3303; the LabConco Model 900/950 burr mill, the most widely used grinder of its type until recently; the Buhler Model ML1204 laboratory burr mill; the Buhler/Miag Model 12285 mill; and the Brabender Schrot mill. All of these mills operate at about 2400 to 3000 rpm, and have one fixed and one rotating burr. Particle size is determined by burr shape and by adjusting the distance between the burrs. Uniformity of particle size and shape when grain is ground using a burr mill is strongly affected by the moisture content of the material. Burr mills usually have an open chamber and the meal does not pass through a screen. A special type of burr mill is the Hobart Model 2040 coffee grinder, whose vertical tungsten steel burrs are tooled in a manner analogous to grooved millstones. Burr mills are very durable and very sensitive to grain texture in terms of the MPS distribution of the resultant meal. They are direct-drive grinders, operating directly at the same speed as the motor, and as a result tend to be noisy.

The second class of grinder is the cutting type. Mills with a cutting action work through a rotating central shaft that carries up to four “knives,” which are really pieces of hardened steel, one end of which has been ground to form a sharp-edged chisel-type knife. The wall of the chamber carries two stationary knives. The sample is chopped into fine particles by the action of the two sets of knives. Cutting action mills usually operate at about the same rpm as burr mills. They operate with a closed chamber in that the meal is forced through a screen. Cutting action mills generally cause less moisture loss than other mills.

The third type of grinder is the hammer mill. Hammer mills consists of a two- to four-armed beater bar that rotates inside a closed chamber, and the meal is forced through a screen on its way to the sample receiver. Hammer mills usually gyrate at about 8,500 to 10,000 rpm. They also tend to be noisy due to the higher rpm and air displacement. They are extremely robust, and the hammer or beater bar usually requires no attention for the life of the grinder. Occasionally screens wear out and the holes of the smaller screens gradually increase in size under heavy use. This is partly a temperature effect. Wear is faster if the screens are continuously hot due to heavy workload. Most hammer mills have rigid beaters, but some have swinging beaters.

Impeller mills are the fourth category of grinder. These consist of a high-speed two-arm impeller whirling at 25,000 to 30,000 rpm in round or oval grinding chambers. The cutting arms are made of hardened steel and the chamber of stainless steel with a plastic cover. A stainless steel liner will greatly prolong the life of the grinder cover. A special type of impeller mill is the Udy Cyclone grinder, in which the impeller is an aluminum vacuum cleaner fan adapted to grind the material by firing it at high speed (13,000 rpm) against an abrasive surface. Unlike other impeller mills, the cyclone grinder forces the meal through a screen, while the air flow passing through a plastic cyclone channels the meal into a receiver jar. Yet another type of impeller mill is the Waring blender. Commonly used to prepare milk shakes in the kitchen, it can serve as a very useful grinder for small samples and also as a blender for emulsifying lumpy slurries. It differs from the cyclone and small impeller mills in that its speed of rotation is much slower. The Waring blender is difficult to clean after use with slurries or oleacious materials.

There are other one-of-a-kind grinders, including the Retsch centrifugal mill, which is the fifth type of grinder. This device carries a turret-shaped impeller surrounded by a circular screen through which the sample is forced into a circular sample receiver chamber. The impeller revolves at either 10,000 or 20,000 rpm, depending on the material to be pulverized. The subject of grinders has been reviewed in connection with grains [12]. The main types of grinder and some examples are given in Table 13.2.

13.5.2 FACTORS AFFECTING GRINDER PERFORMANCE

The main factors affecting grinder performance are summarized in Table 13.3.

Grinder rpm affects MPS, particle size distribution (PSD), particle shape, rate of throughput, temperature increase during grinding, moisture loss, and power consumption. MPS is generally inversely related to rpm, whereas rate of throughput, sample temperature, moisture loss, and power consumption all increase directly as rpm. In the use of the Udy cyclone grinder a reduction in rpm from 13,000 to 12,000 will increase MPS by about 5%, mainly by reducing the amount of very fine particles. The rpm of a motor, and therefore of the grinding element of a grinder is directly related to power input. Power supplies often vary by $\pm 5\%$ during a day due to peak loads and other factors. The corresponding fluctuations in rpm can cause a significant changes in MPS and PSD. This in turn affects diffuse reflectance, bulk density, packing characteristics, moisture loss, and other features of the output of the grinder. This factor becomes more important with high-rpm grinders. Ideally for the most precise performance, grinders, as well as NIR instruments, should work through a voltage regulator. In general, the signal from an NIR or NIT instrument itself on a single sample is more consistent than the characteristics of the powdered materials it is used to analyze.

TABLE 13.2
Main Types of Grinders Available

Grinder type	Grinding system	Typical grinder models
Burr	Vertical, indented burrs	Hobart Models 2040, ^a 2030 ^a
	Elevated graduated burrs	Falling Number (FN) KT 3303 ^b
	Horizontal indented cone	Miag 12285B ^a
	Indented discs	Glen Creston S 80
Cutting action	Four rotating, two stationary blades	Wiley No. 3
	Six pairs of blades	European Brewing Congress (EBC)
	Angled blades	Casella
Hammer mills	Rigid beater cross	Christy/Norris 8-in. FN KT-3100
	Swing-out hammers	Glen Creston DFH 48
	Rigid beater	Labormim Model 1 (Hungary)
Impeller	Two-blade oval chamber	Krups 75
	Two-blade round chamber	Braun KSM-1, Moulinex
Centrifugal		Culatti (Heavy-duty)
	Fan-type impeller with cyclone	U-D Cyclone CSM-1, Tecator CycloTec
	Rotating pins	Retsch ZM-1

^a No longer commercially available, but typical of type.

^b Reference to a specific grinder does not constitute an endorsement.

TABLE 13.3
Factors Affecting Grinder Performance

Factor	Type of grinder affected
RPM	All
Types and design of burrs	Burr mills only
Sample composition (moisture, oil, fiber, etc.)	All ^a
Sample physical texture (hardness)	All
Size of grinding chamber	Cutting-type, hammer
Size and design of cutter knives or beaters	Cutting-type, hammer
Number of cutting edges or beaters	Cutting-type, hammer
Screen aperture	Hammer, impeller, cutting-type, centrifugal
Type of receiving container	Hammer
Air flow	Hammer, Cyclone, CycloTec
Workload	All, especially burr mills
Size of material to be ground, e.g., seeds	Cyclone, CycloTec
Type of material to be ground	All

^a Cyclone and CycloTec not strongly influenced, burr mills strongly affected by moisture or fiber content, high-speed impeller and centrifugal mills not affected by oil content.

Types of burrs also influence both MPS and PSD. Typical of this effect is the relative performance in terms of particle characteristics of ground hard red spring wheat when ground in three different types of burr mills, all with different shapes and sizes of burrs. The Hobart model 2040, the Buhler model ML1204 laboratory grinder, and the LabConco model 900 burr mill gave MPS of 245, 435, and 440 μm , respectively. The Buhler and LabConco meals differ in that the LabConco grinder produced a whole-meal with a much higher proportion of fine particles than the Buhler. Burr shape

also affects durability and throughput. The Hobart 2040 mill with its vertical millstone-shaped burrs had a significantly faster throughput, including sample-to-sample cleanup, than the other two. The more recently introduced Falling Number models KT30 and KT3303 are easy to clean and have high throughput, but are as susceptible as the other burr mills to voltage/rpm fluctuation.

The performances of burr- and cutting-type grinders, and any grinder with relatively low rpm are more susceptible to fluctuations in sample composition, particularly in moisture and oil content. Burr mills usually give poor performance when sample moisture content is above 15%. Similarly, these grinders are not suited to grinding seeds with high oil content. Soybeans, which have 18 to 21% oil, can be ground in burr mills, but cleanup makes the grinding of soybeans very tedious. Grinders with screens are generally not suitable for grinding seeds with oil content of above 5% without cleaning of the grinder between samples.

When oil content climbs to 10% or higher, the oil acts as a pasting agent, the screens rapidly become clogged, and it becomes difficult to grind high oil content seeds without excessive heat buildup and oil extrusion. The Retsch centrifugal grinder is less susceptible to this than other grinders with screens, particularly when grinding canola (rapeseed) or similar oilseeds, and is the most suitable grinder for all types of oilseed (except palm kernels). Small-chambered high-revolution impeller mills are also useful (and much cheaper) for grinding oilseeds. In these mills 20 to 30 g of seeds can be reduced to a fine uniform powder by grinding in up to four 15-s “bursts,” with the partially ground grain being moved or stirred inside the chamber between bursts.

Palm kernels, faba beans, and other large, very hard seeds can be ground by a two-stage system. One convenient method is to pregrind the seeds in a hammer mill fitted with a screen with large (8 mm) round holes. The seeds are ground in a few seconds and emerge as a coarse meal, which can then be ground in an impeller mill (for high oil content seeds) or a cyclone or similar grinder. Forages and highly fibrous materials such as wool or cotton can be ground in mills with a cutting action, such as the Wiley or Culatti mills. The main effects of the fiber are to increase particle length in proportion to width and to decrease bulk density.

Rice presents another factor in grinder performance. While several types of grinders can be used to grind rice, paddy (rice with the hull adhering) is extremely abrasive, and can cause wear of grinder burrs, beaters, impellers, and screen apertures more quickly than any other grain [13].

What is the ideal grinder? The ideal grinder should be capable of grinding a wide range of materials, and giving a fine grind, with consistent PSD with all of them. It should not heat up unduly during grinding many samples, and should be self-cleaning, or at least easy to clean. It should be capable of handling a range of sample sizes, for example from 50 to 300 g or more, or from 1 to 20 g for small grinders. It should be durable, and not require excessive maintenance in the way of replacing burrs, impellers, hammers, screens, etc. It should be capable of grinding without excessive dust or noise, and finally it should be available commercially at a competitive price. MPS, PSD, and, to a certain extent, particle shape are all affected by grinder type, rpm, grinder screen size, feed rate, and composition, with particular emphasis on moisture, oil, and fiber contents. Of the three, moisture content is the most important, and ideally all samples should be at a reasonably uniform moisture content. There is no grinder that is ideal for all commodities, but Table 13.4 indicates some grinders which are best-suited to grinding certain materials.

Although oil content has a very significant influence on particle characteristics, it only affects oilseeds, oil-bearing legumes (soybeans and peanuts), and certain other commodities such as oats, corn, and lupin, wherein oil content can vary from 3 to 10%. Moisture content affects the grinding characteristics of all grains including oilseeds. Grinding of fibrous substances that are high in moisture (15% or higher) results in meals with a higher proportion of longer (larger) particles, because the interaction between cellulosic substances and moisture content has the effect of increasing the proportion and length of longer particles.

To ensure the highest degree of uniformity in particle characteristics, feed rate to the grinder should be regulated. In general, fast feed rates (4 to 5 g/s) cause an increase in particle size, while low feed rates (0.5 to 1 g/s) result in smaller particle size. In wheat, this can amount to as high as

TABLE 13.4
Suggested Grinders for Specific Laboratory or NIR Testing

Workload	Commodity	Grinder
High	Grains, cereals, pulses	KT 3303, U-D Cyclone, ^a CycloTec ^a (Foss/Tecator)
Medium	Grains, cereals, pulses	KT 3100, U-D Cyclone, ^a CycloTec, ^a Retsch, Christy/Norris 8 in., Wiley Intermediate, Glen Mills S 500, others
Medium	Oilseeds	Retsch, Krups 75, Culatti, Moulinex, Braun
Medium	Forages	Christy/Norris 8 in., Cyclone, ^c Wiley No. 4
Prereduction	Forages, hard seeds	Christy/Norris 8 in., ^b Wiley No. 4

^a Fitted with 1.0 mm screen.

^b Fitted with 8 to 10 mm screen.

^c Fitted with forage head.

40 μm in MPS depending on the type of wheat (differences usually greater with hard wheat types). High throughput grinders such as the Udy Cyclone grinder, the Falling Number KT3100, and the Christy/Norris hammer mill should be fitted with a feed rate regulator. Excessive feed rates also overload the grinder, and cause broken driving belts, increased wear of screens and abrasive liners, and other problems.

13.6 BLENDING OF SAMPLES

After grinding, all samples must be thoroughly blended before NIR (or any other) analysis. This is usually done by mixing in the receptacle, or in the sample container. Liquids must also be well blended and free from air bubbles or layers. Layers or strata in liquids are associated with changes in composition and/or changes in refractive index, both of which can affect accuracy and precision. This is particularly important when flow-through cells are being used. The temperature of the cell and the reservoir leading to the cell must be controlled to prevent variation in density and refractive index of the liquids. Slurries and other viscous liquids must also be thoroughly mixed. Ideally liquids should be continuously agitated during scanning to prevent layering. This is particularly important in the case of milk, wherein changes in composition and particle (fat globule) distribution are inevitable.

When substances are sampled in the liquid form and solidify in the sample cell, no stratification or changes in the sample surface should occur. If it does, direct reflectance analysis of such materials can give inaccurate information as to the composition of the material. Slurries of coarse material in a suspending medium should be “ground” by passing the slurry through a high-speed impeller to reduce the size of the particles of the coarse material to improve analytical precision. The simplest case is the direct analysis of solid materials by application of the NIR sensing unit to the surface, or by interactance. Sufficient readings should be taken to ensure adequate precision and representation of the composition, but no blending is involved.

13.7 SAMPLE STORAGE

In most routine analysis situations for quality control, storage of samples is not a factor because samples are either analyzed almost immediately after preparation or, in the case of in-line analysis (analysis carried out directly on the material during and immediately following industrial processing), the sample arrives, is scanned, and leaves the sensing area, and there is no need for storage. When sample storage is necessary, it is essential that the sample presented for analysis be unchanged from that which was taken, in terms of composition and physical form. Sample storage includes storage

of samples before preparation, and after preparation but before analysis. It also includes long- and short-term storage. For storage of samples before preparation, short-term storage usually means up to 2 to 3 weeks, while for long-term storage the term can be a year or more. For samples after preparation including grinding, short-term storage can vary from an hour or two to overnight, while long-term storage should not be more than a week.

The most common changes are changes in temperature, moisture content, volatiles such as low molecular weight fatty acids, other simple acids such as lactic acid, esters and carbonyl compounds, and free fatty acids. Oxidation can also occur. Temperature is a factor that is often overlooked in NIR analysis. Samples ground in high-speed impeller-action grinders can become quite warm during grinding. Materials, such as wheat flour, often reach temperatures as high as 50°C during processing. This should be taken into account during the development of calibrations. In-line analysis may be carried out on the warm (hot?) material, and samples at the full temperature range should be included in the calibration and validation exercises. Flour milling, and many other types of processing in which the products are analyzed in-line, must be considered as a form of sample preparation, since the milling operation is the last thing that happens to the sample before analysis.

Changes in moisture content during sample preparation are particularly dependent on the original moisture content of the sample. The laws of physics are inviolate and moisture exchange with the surrounding atmosphere is inevitable if the moisture content of the sample is not in equilibrium with the relative humidity of the atmosphere surrounding the sample. In dry atmospheres, grains and forages can lose moisture until they reach levels as low as 5 to 8%. Low moisture content can affect the way in which substances such as grains reduce on grinding. This, in turn, can affect the particle characteristics and their diffuse reflectance, and can cause discrepancies in NIR analysis. Conversely, material can also absorb moisture to the extent that mold may develop. This can occur in cold storage. Opening the door to remove or add samples causes warm air to enter the cold storage room. This warm air will carry more moisture, which will leave the atmosphere in the cold room, and become absorbed by any samples exposed to the air. Samples stored in cold storage should be protected by storage in heavy plastic bags, or in moisture-tight metal or plastic containers.

Material with very high moisture content, in other words, above 15%, must be refrigerated if storage periods of more than 24 h are anticipated, to prevent the development of molds or bacterial activity, but ideally it should be analyzed directly after sampling. Silages and fresh grass or leafy forages should be dried before analysis and the moisture content determined, with suitable precautions to avoid loss of volatiles so that the protein content or that of other constituents can be reported accurately. An error of 1 or 2% in reported moisture content will affect the accuracy of determination of all other constituents and, as a consequence, moisture is probably the most important constituent of all materials. Meats and fish, and fresh vegetables and fruit all have moisture contents of 70% or higher, and should be analyzed by NIR or reference methods directly after sampling.

Materials of lower moisture content (up to 15%) should be stored in metal or plastic containers and sealed or covered with tightly fitting lids to prevent moisture exchange. Samples should be carefully blended before storage to facilitate accurate subsampling before analysis. For long-term storage, ideally all material should be stored under refrigerated conditions. Materials with high oil content are the most difficult to store because they are prone to deterioration by moisture loss which, in a closed container, can stimulate the development of fungi, and also deterioration of the oil. Deterioration includes oxidative rancidity, increases in free fatty acids and other factors, and also depends on the composition of the oil. For example, oats with an oil content of only about 5% are very susceptible to rancidity. Ground oilseeds undergo changes caused by oxidation and agglomeration that affect diffuse reflectance. It is advisable to analyze oilseeds such as rapeseed, soybean, flaxseed, peanuts, etc., as soon as practicable after grinding.

The amount of material stored can also be a factor in the efficiency of storage. The amount necessary for complete analysis should be calculated, then twice the amount stored (in case of possible reanalysis) plus about 10% to allow for spillage. Storage of quantities substantially greater than needed ties up containers and storage space, which results in extra expense and may reduce the

efficiency of storage. Plant material, such as straws or grasses, should not be stored in the “long” state, but should be subsampled, dried, and ground before storage to reduce the size of containers and storage space.

Storage of samples after preparation for analysis should be in containers that facilitate mixing before NIR analysis, or weighing for reference analysis. The containers should also enable taping to reduce moisture loss. Two- or four-ounce ointment tins are ideal for sample storage, because they hold enough of any ground material for NIR analysis, the material can be easily mixed, and the tins are straight-sided and easy to tape. An array of tight-lidded plastic containers are also available for this type of storage. In areas where the relative humidity is naturally high, ground-dried materials may gain moisture, and these precautions are equally applicable. Storage of ground samples in plastic containers may cause some difficulties in analysis due to static electricity, especially in dry climates. Fibrous materials in particular often become orientated, and stick to spatulas and the walls of the containers.

13.8 SAMPLE CELLS AND SAMPLE PRESENTATION

This section refers to the actual presentation of the sample to the NIR instrument. First, a word on sample presentation cells. Most NIR instruments have some sort of cell by which the sample is presented to the instrument. As a result of interaction between customers and instrument manufacturers for specific applications, sample presentation cells occur in a multitude of shapes and sizes, although most companies retain the same sample cell for their off-the-shelf NIR models. A successful sample cell should be easy to fill without stratification, easy to clean, and should provide a sample layer deep enough to ensure diffuse reflectance from the sample only, without the possibility of the signal reaching the detector consisting partly of reflectance from the cell cover.

Very small samples can be analyzed by placing a ceramic (or other) standard between the sample and the cell cover. In this way, spectra have been recorded from samples as small as 10 μ l of liquid and a few milligrams of solid material. It is possible to emphasize a spectral anomaly (Woods anomaly) that occurs around 1500 nm if the surface presented to the instrument is not uniform. This will appear as a band of varying intensity.

The instruments have no sample cell, and the sample is accessed to the optical system through an impeller. Subsequently, it has been shown that complications can arise with this method of sample access. For example, some materials, such as high-moisture grains, may not flow evenly through the instrument. This has been overcome by some NIT instrument companies by developing a sample transport system comprising a cell and a system for accessing the cell to the optics. Cells of different widths provide path lengths to suit different types of grains. Modern NIT instruments automatically adjust the path length to suit each commodity.

Loading of sample cells may appear to be the easiest step in NIR analysis. In practice, it is one of the most critical, in that the bottom of the layer of sample is what the instrument actually senses. Stratification during filling the cell can affect accuracy and precision. All the care taken in sampling and sample preparation is wasted if insufficient attention is paid to loading the cell. A recent study of replicated NIR analysis of a series of wheat cargoes showed that cell loading was a bigger source of variance than grinding [14]. The cell-loading error can be determined by loading the cell several times with the same ground sample, and rotating it through 180°. Readings are then taken without rotating the sample, so that the instrument is simply rescanning the sample. This tests the instrument error vs. error induced by cell loading and stratification. The sample must be thoroughly remixed between loadings.

The only guideline to efficient cell loading is that of care and attention to ensure even distribution of the sample against the cell window. Use of several sample cells can speed up analysis of large numbers of samples. Small differences may occur between quartz or glass cell windows, and reproducibility of results between cells should be checked. Some glass windows show an absorber near 2180 nm, which can interfere with calibrations for protein and other nitrogenous constituents.

The thickness of glass windows is more variable than is that of quartz, which causes window-to-window variance. If more than one cell is to be used with glass windows, intercell agreement should be checked before use.

Filling the cell with the same amount is also important. This can be attained by striking a level sample surface before closing the cell. Overfilling or underfilling the cell changes the degree of compaction at the surface. This changes the diffuse reflectance and hence the NIR results. Overfilling can also cause extrusion of oil, in the case of oilseeds [15], and for very highly oleaginous samples an open-top cell with no cover may give the best results. The Percon Inframatic NIR instrument presents a special case in that there is no cell as such. The degree of compaction of the sample in the Inframatic sample compartment is important. Undercompaction causes inaccuracies, while overcompaction may force the sample compartment open.

13.8.1 CELL CLEANUP

This is yet another factor in accurate and consistent NIR analysis. It is affected by the composition and degree of fineness of the material to be analyzed, by the material of which the cell and spatula are made, by the design of the cell, by the moisture content of the analyte and of the atmosphere, and by other factors, including the clothes of the operator, which can introduce static electricity. In general, the higher the bulk density of a grain, the finer will be its particle size on grinding particularly in an impeller or hammer mill. Peas, chickpeas, lentils, and beans give meal with the highest bulk density (Table 13.5), followed by cereals, oilseeds, and forages.

Sample cells are more difficult to free from very dense meals and flours, and it is usually necessary to wipe the cell cap and glass/quartz cover surface with soft tissues. NIR analysis of oilseeds unavoidably results in extrusion of some oil onto the surface of the cell window. Between samples of oilseeds, butters, or any other high-fat or high-oil material, the window should be cleaned with alcohol, then water to remove the alcohol, followed by careful drying, because both alcohol and water have characteristic bands in the NIR region. Fibrous materials may cause difficulties in cell cleanup due to static electricity. Cleanup of flow-through cells can be achieved by washing with a liquid such as water, or (more efficiently) by allowing the next sample to pass through the cell for several seconds before taking the next readings. Viscous liquids and slurries are similar to oilseeds in that they require washing and drying of cells and covers between samples.

TABLE 13.5
Bulk Densities of Some Common Powdered
Plant Materials

Commodity	Bulk density (g/cm ³)
Faba bean (<i>Vicia faba</i>)	0.639
Lentil	0.542
Wheat flour (hard wheat)	0.514
Corn (maize)	0.506
Wheat whole-meal (hard)	0.495
Soybean hulls	0.489
Barley	0.390
Undefatted soybean meal	0.360
Wheat bran	0.349
Oats	0.345
Defatted rapeseed meal	0.338
Dried alfalfa meal	0.240
Barley straw (mature)	0.142

13.8.2 STATIC ELECTRICITY

Static electricity causes orientation of the powdered sample at the cell cover surface, which can seriously affect precision. This is most apparent when the cell window is of quartz rather than glass, and when the moisture content of the material is very low. Low atmospheric relative humidity and repeated cleaning of quartz cell windows by rubbing with tissues or brushing also increase static electricity. Other factors that affect static electricity include the presence of carpeting made of synthetic materials, plastic spatulas, nylon brushes, and a high proportion of nylon or other synthetic textile in the clothing of the operator (including underwear). Storage of ground materials, particularly fibrous materials in plastic bags, may cause so much static electricity that it is difficult to get the sample out of the bag and into the cell.

13.8.3 UNDETECTED MOISTURE

Oilseeds and other seeds, the results of which are reported on a moisture-free basis, are usually dried before analysis by reference methods, because the moisture interferes with the determination of oil by extraction. For the most accurate results, the seeds should be dried in a vacuum oven after grinding, then analyzed. If the seeds are dried whole, then ground for analysis, the ground material will absorb water rapidly, and may contain 2 to 4% moisture by the time it is analyzed. This undetected moisture can induce significant errors to the reporting of reference results for oil or protein [16]. Figure 13.8 illustrates the difference in the moisture band at 1930 nm for soybeans dried before and after grinding. The moisture content of the six samples of dried soybean seed varied up to 3 to 4% enough to cause errors of about 2% in protein content. The reason for the absorption of water in the presence of very hygroscopic cellulosic fiber in the oilseed testa-pericarp. If the oilseeds are to be analyzed by NIR there is no need for drying the samples before analysis. Moisture must then be determined to enable reporting of the oil, protein, etc., results on a moisture-free basis.

Agglomeration of ground material after grinding is a source of error in oilseed analysis by NIR. Agglomeration of particles affects particle behavior and, consequently, diffuse reflectance [17]. Oil acts as a pasting agent, and the degree and speed of agglomeration is positively related to the oil

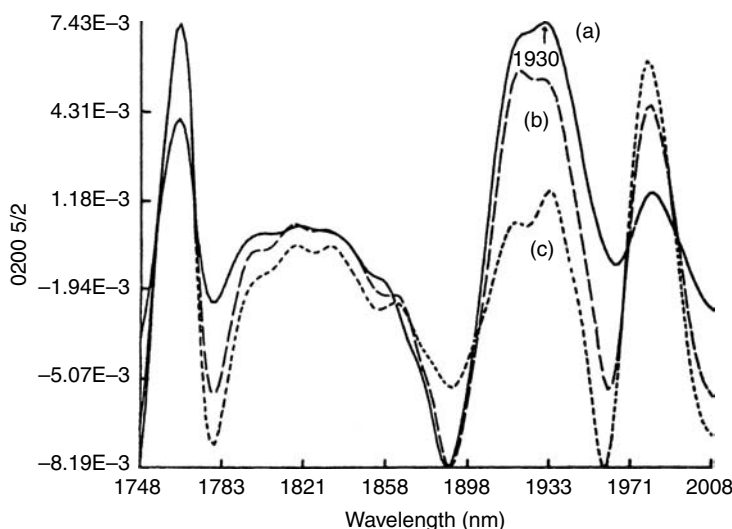


FIGURE 13.8 Inverted second-derivative NIR spectra of soybeans dried in vacuum oven at 100°C overnight. (a) Not dried at all; (b) seeds dried, then ground; and (c) seeds ground, then dried. The water bands in the 1930 nm area in spectrum (c) are significantly lower than those in spectrum (b).

content. Samples should be analyzed by NIR or used in calibration within 1 or 2 days of grinding to avoid this source of error, and it is important that they be prepared in the same way and stored for the same time as the samples used in calibration. The introduction of NIT instruments capable of accurate analysis of whole oilseeds will likely prove to be very valuable in oilseed analysis, because NIT eliminates grinding, oil extrusion, moisture, and agglomeration errors.

Fruits and vegetables present a special case. First, they have a very high moisture content, and second, they vary widely in size and color. They can be analyzed by transmittance [18] or by reflectance. For analysis by reflectance, a good technique is to take a slice of the individual fruits or vegetables, and read the surface directly. It is also possible to pass the material through a high-speed blender to produce a slurry for presentation to the NIR instrument. One advantage of the slurry approach is that several pieces of fruit or vegetable can be blended to reduce sampling error. Meats and fish, which also have high water contents, can be analyzed directly using a fiber-optics attachment or after mincing in a mincer such as a sausage-maker, which produces a very finely minced product similar to sausage meat.

Some types of high moisture samples can be scanned and analyzed directly in plastic bags. This will effectively preserve the moisture status of the sample, as well as avoiding loss of volatiles. This technique is particularly applicable to samples of materials with obnoxious odors, such as manures, or materials that may give off harmful vapors. It is also useful for quality control analysis of packaged materials, because the materials can be analyzed without removing them from their package. An empty plastic bag, or portion of a bag, can be used as reference. The instrument software will subtract the reference from the sample spectrum, which should eliminate spurious absorbance bands arising from the plastic.

Liquid samples can also be scanned directly in scintillation vials. Some instruments offer special attachments for this type of presentation, such as the Rapid Content Analyzer attachment for the Foss/NIRSystems Models 6500 and 5000.

Instrument companies such as Bran + Luebbe, Foss/NIRSystems, and LT Industries have designed and perfected an extremely diverse array of sample presentation systems for application with specific materials. This include in-line systems for continuous monitoring of materials during industrial processing. These may be used in conjunction with fiber-optics attachments that carry optical data to spectrophotometers, which can then be installed in a location where they will not be exposed to the rigorous conditions that may persist at the actual processing site. There are many examples of installations of custom-designed sample presentation systems in industry. In the grain industry, whole-grain analyzers have minimized errors incurred by sample presentation.

13.9 SAMPLE SELECTION

In the case of NIR instruments, sample selection is really a factor only in calibrations and their validation, because the operator has no choice in the samples presented for routine analysis. Calibration has been a bone of contention ever since the dawn of the NIR era. The most serious criticism of NIR technology is that the instruments require separate calibrations for different commodities for the same constituent, as well as for different constituents of the same commodity. There are two main methods of calibration. The first calls for accumulation of sufficient samples with reference analyses to enable generation of calibration and verification files, followed by selection of samples to fill these files. The second method involves selection of samples strictly on the basis of spectral characteristics, followed by reference analysis on only the relatively small number of samples that display the most comprehensive variance in optical data. These two methods are referred to respectively as conventional and spectral methods of sample selection in this chapter.

Conventional methods of sample selection involve identification of all sources of variance likely to be encountered in future analysis including sample source (growing location and season) and the range of composition in constituents or parameters to be tested. The most important factors that affect the behavior of solid samples are chemical composition, physical texture, bulk density,

processing method, temperature, and color. The first three factors are mainly responsible for packing characteristics in the sample cell, which affect the diffuse reflectance from the surface. Color affects the gross reflectance of light energy from the surface and can introduce biases, particularly at lower wavelengths.

Chemical composition affects the sample in two ways. First, a composition factor. The amount of water, oil, and fiber present all affect the way a sample grinds. Generally, the higher the moisture, the coarser the MPS of the ground material. High moisture can also affect the shape of particles, which become “flaky” at high moisture levels. Accurate determination of moisture becomes very critical with very high-moisture substances, because errors in the moisture test will affect the accuracy of reporting the results of all other analyses. It can also cause materials to pack more densely. Oil acts as a pasting agent and has a very important influence on packing characteristics. Materials with high oil content will form a densely packed surface with lower diffuse reflectance than low oil content substances. The presence of high cellulosic constituents generates long narrow particles, which pack differently than materials lower in fiber, such as ground grains.

Second, the presence of high concentrations of constituents, which absorb in the same area, can cause the selection of wavelengths in alternate areas for one or more constituents. For example, for highly fibrous oilseeds, wavelengths selected for the determination of oil may be in the 12- or 1700-nm areas rather than in the “traditional” 2300-nm area.

It is important to select samples with uniform distribution with respect to the range of constituent or constituents to be determined. Selection of samples with Gaussian distribution of samples will cause the results of subsequent analysis to regress toward the mean. This phenomenon, called the Dunne effect [19], is most pronounced in sample sets with very large variance and lower coefficients of correlation between NIR and reference analyses. A minimum of 100 samples should be used in a calibration. If the range in composition is, for example, from 9 to 18%, this means that about 10 samples each should be included in the range 9–10, 10–11, etc., up to 17–18%. Operating manuals provided by the instrument manufacturers give detailed instructions for sample selection.

The biggest drawback with uniform selection of samples is that it is usually necessary to analyze a lot of samples in order to identify sufficient samples with required composition toward the extremes. Furthermore, it may be necessary to analyze an even greater number of samples if several constituents are to be determined. In practice, this can best be achieved by storage of analyzed samples prior to calibration, and carrying out the calibration when sufficient samples have been accumulated.

13.9.1 SPECTRAL SAMPLE SELECTION

During the past 15 years an alternative method for sample selection, based on spectral characteristics, has been introduced [20,21]. The procedure calls for recording the spectra, then selecting the samples likely to provide the best calibration for a given constituent based solely on spectral variance. The selected samples are then submitted for reference analysis. This method significantly reduces the number of samples to be analyzed by the reference methods and, therefore, the expense of the laboratory testing. Drawbacks to this methods are (a) a large number of samples must be assembled and scanned in order to provide spectra displaying maximum variance. All sources of variance (composition, growing environment, etc.) must still be identified, and samples assembled to provide spectral variance from all sources, (b) different sample sets may be necessary for different constituents, (c) modern NIR technology has expanded into the determination of minor constituents such as antinutritional factors in food legumes and cereals, and spectra often do not differ very much in the wavelength areas used in determination of these constituents. Use of partial least-squares (PLS) regression can overcome this hurdle.

A fourth factor is that physical texture has a very significant effect on particle size and shape, and therefore on spectral characteristics. Variance in spectra may be caused by differences in particle size, rather than to chemical composition. As a result, sample sets may be selected with unsatisfactory distribution in composition across the anticipated range for the constituent to be determined.

Alternatively, a very small sample set can be selected, which may not lead to a stable calibration. Modern methods of spectral selection incorporate principal component analysis [22], which uses all wavelengths available. This procedure improves spectral sample selection. Another method of improving the integrity of spectral sample selection is to derivatize the $\log(1/R)$ data before selection. Derivatization reduces variance caused by particle size and, as a result, differences between spectra are more likely to be due to chemical than physical differences.

Modern NIR software, such as WINISI offers excellent methods for spectral sample selection, and is another procedure, which is likely to become much more widely used in the future. The philosophy is that when two samples have spectra that are very closely similar there is only a need for one of them. In this way, Gaussian distribution of calibration samples and the Dunne effect are largely eliminated, because many samples will be spectrally similar, and only a relatively small proportion of a population will be selected. This will also reduce the need for expensive reference analysis.

Probably the best advice to give to an aspiring NIR user is that, when data are available sample selection on the basis of chemical or physicochemical analysis is the safest route to take. Agricultural and food commodities are very complicated substances, with oil, protein, moisture, and cellulosic components present in various degrees and in substances with very big differences in physical texture and functionality. The scope for interactions is almost infinite. For applications involving simply moisture or a single chemical constituent in a plastic, textile, pharmaceutical, or some other industrial product, spectral selection may be preferable. Finally, for some industrial applications where the product is simply too big to sample and the NIR instrument has to be brought to the product (sample) surface, or for in-line analysis it may be necessary to prepare samples synthetically by preparing small batches of the product with deliberately varying composition.

13.10 SUMMARY

Sample selection, sampling, sample preparation, and sample presentation to the instrument are fundamental to accurate and precise testing by both NIR and reference methods. This chapter summarizes the steps involved in all three operations. Of these, sample preparation is arguably the area from which most errors arise, because it is hard to prepare a sample without changing it in some small way from its original form — the form in which it will be used or consumed. Sample preparation includes documentation, blending, subsampling, removal of unwanted material, grinding, and reblending after grinding and storage. Sample presentation to the instrument also ranks high on the list of things to which attention must be paid, although most of the responsibility for this lies with the engineering of the instrument. In many cases improvements in the statistics of the effectiveness of a NIR calibration lie in improvements in one or more of the factors involved in the sample itself, its preparation, and its presentation to the instrument.

Particle size, moisture content, and temperature, the three main factors identified early in the NIR era, are all intrinsic to effective sample preparation and presentation.

REFERENCES

1. P. C. Williams, K. H. Norris, C. W. Gehrke, and K. Bernstein, Comparison of near-infrared methods for measuring protein and moisture in wheat. *Cereal Foods World*, 28: 149–152, 1983.
2. J. M. Conway, K. H. Norris, and C. E. Bodwell, A new approach for the estimation of body composition: infrared interactance. *The American Journal of Clinical Nutrition*, 40: 1123–1130, 1984.
3. P. E. Parker, G. R. Bauwin, and H. L. Ryan, Sampling, inspection and grading of grain, in *Storage of Cereal Grains and Their Products*, 3rd edn. (C. M. Christensen, ed.), The American Association of Cereal Chemists, St. Paul, MN, 1982, p. 10.
4. P. C. Williams, Commercial near-infrared reflectance analyzers, in *Near-infrared Technology in the Agricultural and Food Industries* (P. Williams and K. Norris, eds.), The American Association of Cereal Chemists, St. Paul, MN, 1987, p. 111.

5. J. F. Lockwood, The Screenroom, in *Flour Milling*, 4th edn. (J. F. Lockwood, ed.), The Northern Publishing Co., Ltd., Liverpool, England, 1960, pp. 209–219.
6. L.-A. Appelleqvist and B. Löf, Post-harvest handling and storage of rapeseed, in *Rapeseed, Cultivation, Composition, Processing, Utilization* (L.-A. Appelleqvist and R. Ohlson, eds.), Elsevier Press, Amsterdam, Holland, 1972, pp. 60–100.
7. R. C. Brook and G. H. Foster, Drying, cleaning and conditioning, in *Handbook of Transportation and Marketing in Agriculture, Vol. II. Field Crops* (E. E. Finney, Jr., ed.), CRC Press Inc., Boca Raton, FL, 1981, pp. 63–110.
8. F. W. Bakker-Arkema, R. C. Brook, and L. E. Lerew, Cereal grain drying, in *Advances in Cereal Science and Technology, Vol. II* (Y. Pomeranz, ed.), The American Association of Cereal Chemists, St. Paul, MN, 1978, pp. 1–90.
9. G. H. Foster, Drying cereal grains, in *Storage of Cereal Grains and Their Products*, 3rd edn. (C. M. Christensen, ed.), The American Association of Cereal Chemists, St. Paul, MN, 1982, p. 79.
10. G. C. Galletti and R. Piccaglia, Water determination in silages by Karl Fischer titration. *Journal of the Science of Food and Agriculture*, 43: 1–7, 1988.
11. P. C. Williams and B. N. Thompson, Influence of wholemeal granularity on the analysis of HRS wheat for protein and moisture by near-infrared reflectance spectroscopy (NIR). *Cereal Chemistry*, 55: 1014–1037, 1978.
12. P. C. Williams, A study of grinders used for sample preparation in laboratory analysis of grains. *Cereal Foods World*, 29: 770–775, 1984.
13. A. B. Blakeney, personal communication, 1998.
14. W. R. Hruschka, Data analysis: Wavelength selection methods, in *Near-infrared Technology in the Agricultural and Food Industries* (P. Williams and K. Norris, eds.), The American Association of Cereal Chemists, St. Paul, MN, 1987, p. 41.
15. P. C. Williams, Variables affecting near-infrared spectroscopic analysis, in *Near-infrared Technology in the Agricultural and Food Industries* (P. Williams and K. Norris, eds.), The American Association of Cereal Chemists, St. Paul, MN, 1987, p. 158.
16. P. C. Williams and K. H. Norris, Qualitative applications of near-infrared reflectance spectroscopy, in *Near-infrared Technology in the Agricultural and Food Industries* (P. Williams and K. Norris, eds.), The American Association of Cereal Chemists, St. Paul, MN, 1987, p. 241.
17. J. Panford, personal communication, 1986.
18. G. S. Birth, W. C. Turley, and S. Kays, Non-destructive determination of soluble solids in onions. Pacific Scientific, Gardner/Neotec Instruments Div., Tech. Paper No. 4036.
19. W. Dunne and J. A. Anderson, A system for segregating Canadian wheat into sub-grades of guaranteed protein content. *Canadian Journal of Plant Science*, 56: 433–450, 1976.
20. D. E. Honigs, G. M. Hieftje, H. L. Mark, and T. B. Hirschfeld, Unique sample selection via near-infrared spectral subtraction. *Analytical Chemistry*, 57: 2299–2303, 1985.
21. H. L. Mark and D. Tunnell, Quantitative near-infrared reflectance analysis using Mahalanobis distances, *Analytical Chemistry*, 57: 1449–1456, 1985.
22. H. Martens and T. Naes, Multivariate calibration by data compression, in *Near-infrared Technology in the Agricultural and Food Industries* (P. Williams and K. Norris, eds.), The American Association of Cereal Chemists, St. Paul, MN, 1987, p. 64.

14 Indicator Variables: How They May Save Time and Money in NIR Analysis

Donald A. Burns

CONTENTS

14.1 Background	297
14.2 The Problem.....	298
14.3 Procedure.....	299
14.4 Interpretation.....	299
14.5 Further Data Manipulation	301
14.6 The Solution.....	303
14.7 What about More than Two Levels?	303
14.7.1 Antibiotic.....	303
14.7.2 Paper Product	304
14.8 Summary	305
Further Reading	306

14.1 BACKGROUND

In near-infrared spectroscopy (NIRS), as we have seen in earlier chapters, one *teaches* the instrument/computer system what to look for in a given type of sample, then expects the hardware/software combination to produce valid answers when it is presented with unknown samples of the same type. But what kind of answers are we seeking? Usually, they are *quantitative* answers, in other words, how much of substance XYZ is present in the sample. Occasionally we use what is called *discriminant analysis* to obtain a *qualitative* answer, in other words, is this sample A, B, C, or none of these?

But there is a third kind of answer that we ought to consider, and it is based on information (data) that we sometimes have but generally ignore. The software provided by most instrument manufacturers will accommodate this “excess” information, and it behooves us to know that advantages accrue from using it. Let us take it from the beginning.

From earlier chapters we now know the big role played by *chemometrics* in NIRS. An early (and still popular) type of data treatment is *multiple linear regression* (MLR), and one way to express the equation generated by its program is

$$\% \text{ Conc.} = F(1) \times \text{Abs}(W1) + F(2) \times \text{Abs}(W2) + \cdots + F(N) \times \text{Abs}(WN) + F(o)$$

where $F(1), \dots, F(N)$ are “weighting factors” and $\text{Abs}(W1), \dots, \text{Abs}(WN)$ are absorbances at specified wavelengths. $F(o)$ is an offset (bias) and essentially moves the calibration line up or down.

The same equation can be written this way:

$$\% \text{ Conc.} = \sum_{i=1}^N [F(i) \times \text{Abs}(Wi)] + F(o)$$

What we're going to show in this chapter is how an additional term in this equation can often produce better answers. The modified equation can be written this way:

$$\% \text{ Conc.} = \sum_{i=1}^N [F(i) \times \text{Abs}(Wi)] + F(o) + [G \times IV]$$

where the last bracketed term is a second offset/bias that will be zero when the indicator variable (IV) = zero, but will contribute to the value when $IV = 1$. As we shall see, IV can only have these two values, 0 or 1. In other words, we shall merely indicate the presence or absence of a factor rather than assign a continuum of values to it.

A typical calibration for an NIR method might involve the following:

- Thirty samples whose absorbances have been obtained by scanning over the wavelength range 1100 to 2500 nm.
- Varying levels of some constituent, previously determined by an acceptable laboratory (reference) method.

If this is the maximum amount of information that can be provided, then the system must work with what it has. But more often than not there are additional data that could be supplied and are not, because they aren't deemed to be important at the time. Examples of such additional data might be

- Which one of two chemists prepared each sample?
- Which one of three suppliers provided the starting material?
- Which samples were air dried at room temperature, and which ones were heated to 55°C in an oven?

This seemingly unimportant (irrelevant?) information can be included in the calibration procedure, and it just might lead to our knowing more about the samples than we thought possible. For instance, does it really make any difference *who* prepared the samples? Does the *source* of the starting material affect the answer? Does drying temperature have *any* effect on the final product? Answers to such questions as these can be obtained using so-called indicator variables in the calibration procedure. In this chapter we'll learn how this is accomplished by following the course of a real calibration. Because of the proprietary nature of some of the materials, constituent names have been changed and values have been scaled to maintain confidentiality.

14.2 THE PROBLEM

An electrically conducting polymer is made with a stabilizer that is present at levels ranging from 0 to 45 units. There are six additional differences in the samples, and each of these parameters can be at either of two levels, identified as follows: operator, particle size, vendor, solvent, pressure, and temperature. The questions to be answered are

1. Can NIRS be used to analyze this polymer for the stabilizer?
2. Are any of the six other parameters important (significant)?

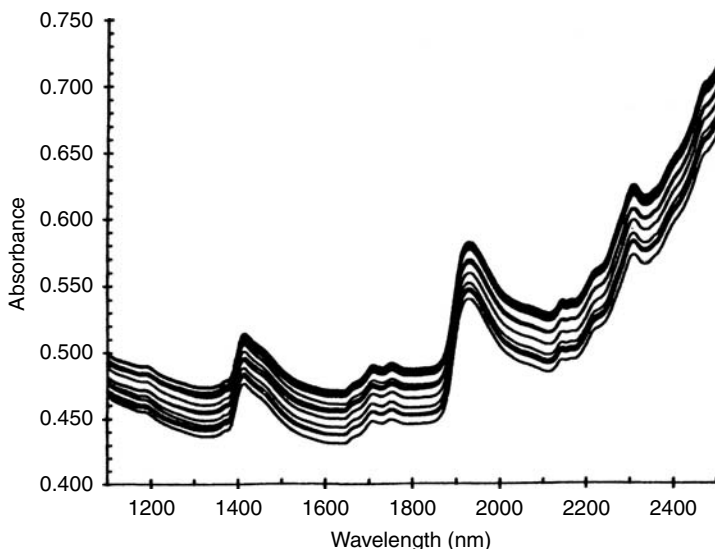


FIGURE 14.1 Family of tracings of 16 duplicate polymer samples.

While we generally settle for an answer to #1, we would be well advised to consider #2 (and perhaps reap such benefits as higher accuracy, better precision, and/or simpler sample preparation).

14.3 PROCEDURE

For purposes of illustration we will take 32 samples (16 run in duplicate). The tracings of the entire set of samples are shown in Figure 14.1. It is immediately apparent that (a) a spectrum exists, so analysis by NIR is probably possible and (b) there are no obvious wavelengths where major differences seem to occur. The full dataset is shown in Table 14.1, and the columns beyond the first two require further explanation.

As the term *indicator variable* might imply, we will “indicate” (with a 1 or a 0) the presence or absence of a given parameter. In the case of two operators/analysts/technicians, the sample was either handled by one specified person or it was not. Thus, operator (A) could be represented by a 1 in the first column and the alternative operator (B) would then be represented by a 0. Two different particle sizes of some variable would be similarly represented, for example, the smaller size with a 0 and the larger size with a 1. We will address the case of more than two levels/operators/sizes/etc. later. For an initial understanding, the case of two alternatives is sufficient.

Two search strategies were available: a traditional *step-up* search and the more powerful *all-possible-combinations* (APC) search. The latter was selected because of its capability of handling indicator variables. The initial search was done over the wavelength range 1100 to 2416 nm and used all available data. Since the single analyte plus the six possible additional parameters is roughly equivalent to seven constituents, the search required several hours to produce its first equation. The parameters and statistics are shown in Table 14.2.

14.4 INTERPRETATION

The F value is >20 , so the equation is reasonable (if not ideal in this real-world case). This statistic is a measure of our confidence in the regression, and it generally indicates how well the calibration line will predict answers. The F value increases as (a) we add more samples, (b) we eliminate wavelengths with little or no “predictive power,” (c) the correlation coefficient approaches unity,

TABLE 14.1
Chemlist of Data File (duplicates not shown). File ID: Polymer
Electrically Conducting

Sample	Stabilizer	Operator	Particle size	Vendor	Solvent	Pres.	Temp.
1	3	0	1	1	1	1	0
2	42	1	0	1	0	1	0
3	24	0	1	0	0	0	1
4	6	0	1	1	0	0	1
5	15	1	0	0	1	0	1
6	9	1	0	1	1	0	1
7	12	1	0	0	0	1	0
8	21	0	1	0	1	1	0
9	0	0	0	1	0	0	0
10	39	0	0	1	1	1	1
11	27	1	1	1	1	0	0
12	45	1	1	0	1	0	0
13	30	0	0	0	0	0	0
14	18	1	1	1	0	1	1
15	33	0	0	0	1	1	1
16	36	1	1	0	0	1	1

TABLE 14.2
Regression for Stabilizer (Range: 0 to 45)

Wavelength	Factor	Regr. coef.	<i>t</i> Value
1226		3,899	5.34
1660		−35,376	−7.39
1758		31,957	7.62
Intercept		−696	
	Operator	−2.33	−1.35
	Particle size	1.61	1.04
	Vendor	13.7	5.86
	Solvent	12.8	6.92
	Pressure	0.53	0.31
	Temperature	2.74	1.74

Correlation coefficient = .965; SEE = 4.38; *F* value = 33.

and (d) we delete true outliers. The initial calibration line is reproduced as Figure 14.2. It reflects the correlation coefficient of .965, and it suggests the existence of (perhaps) one outlier. Of the three wavelengths chosen, one (1226 nm) is an order of magnitude less important than the other two. This is apparent by comparing their regression coefficients. However, the *t* value (Student's *t*-test) was sufficiently high (>2.5) to warrant keeping this wavelength.

An examination of the *t* values for each of the six factors (parameters) reveals that vendor and solvent are truly significant. It is also noteworthy that their contributions could be 13.7 and 12.8 units, respectively, to the stabilizer (which can be predicted with an error of only 4.38 units). Ignoring these contributions (about $3 \times$ the system error) could result in poor predictions.

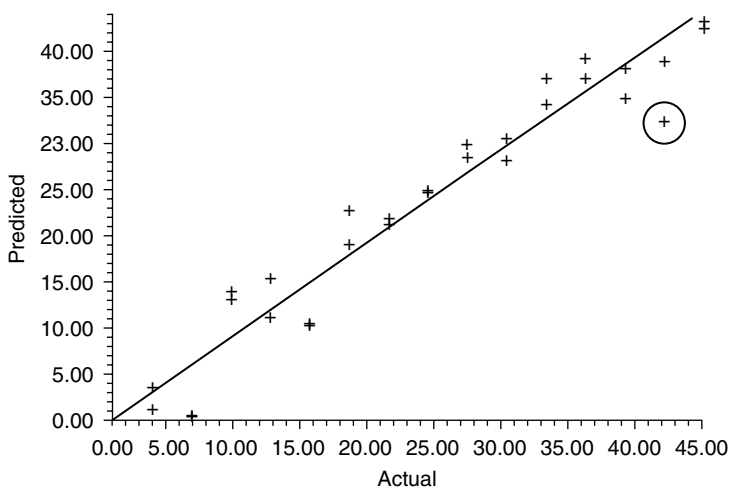


FIGURE 14.2 Initial regression of polymer samples including suspected outlier.

14.5 FURTHER DATA MANIPULATION

On the basis of the above interpretation of this first regression, a second one was done with two differences: (a) The same three wavelengths were “forced” into the equation (to avoid waiting several more hours for a new calibration) and (b) the outlier (which was flagged in the printout) was removed from the input data. The consequence of deleting this outlier is shown in the following table:

	Full dataset	One sample removed
<i>F</i> value	33	42
SEE	4.38	3.78
Corr. coef.	.965	.973

The 27% increase in the *F* value confirms that the deleted sample was indeed an outlier. Likewise, the reduction in the error means the decision to remove it was a valid one. The correlation coefficient would be expected to rise along with an improvement in *F* value.

The next regression was to evaluate the consequence of deleting those parameters with low *t* values. Accordingly, operator, particle size, pressure, and temperature were eliminated (along with the outlier) for the next equation:

	All six parameters	Only two parameters
<i>F</i> value	42	57
SEE	3.78	4.29
Corr. coef.	.973	.959

Clearly, the equation has become more “robust” (*F* value rose another 36%), even though we have lost a little in the other two statistics. The trade-off may be a good one and the elimination of four relatively unimportant parameters could be a time saver in sample preparation. We are now fairly sure, for example, that there is no significant difference between the two operators, that particle size does not seem to influence the ability of the equation to predict stabilizer level, etc. This knowledge is very likely to simplify sample handling.

TABLE 14.3
Regression for Stabilizer (Range: 0 to 45)
Equation (14.4)

Wavelength	Factor	Regr. coef.	<i>t</i> Value
1216		3,526	5.45
1672		−38,001	−7.46
1756		35,067	7.66
Intercept		−738	
	Vendor	14.0	6.25
	Solvent	13.3	7.20

Correlation coefficient = .959; SEE = 4.37; *F* value = 59.

One last regression could serve to fine-tune the procedure by decreasing the wavelength interval from 14 to 6 nm to see if a better equation can be generated by this higher resolution data. To a very slight degree it can be improved, as shown below and in Table 14.3:

	14-nm spacing	6-nm spacing
<i>F</i> value	57	59
SEE	4.29	4.37
Corr. coef.	.959	.959
Wavelength		
#1	1226	1212
#2	1660	1672
#3	1758	1756

A slight change in wavelengths occurred, along with an improvement in the *F* value. The increase in the standard error of estimate (SEE) is a result of including the sample that had been rejected from the previous calibration. With this new combination of wavelengths, all samples could be used (there being no justification to declare one or more to be outliers). This final equation, then, is significantly better than the first one, as seen by the steady improvement in “robustness” without serious deterioration of other statistics from intermediate values. The various values are compiled in the following table for convenient comparison:

	Equation			
	(14.1)	(14.2)	(14.3)	(14.4)
<i>F</i> value	33	42	57	59
SEE	4.38	3.78	4.29	4.37
Corr. coef.	.965	.973	.959	.959
	All data	Outlier removed	Four of six parameters out	Fine-tuned

Could anything more be done to improve the equation? Possibly, by searching with a still smaller wavelength increment, by using a larger teaching set of samples, by reinstating one of more of the four deleted parameters (which, admittedly, have a small contribution), and perhaps by adding additional wavelengths.

14.6 THE SOLUTION

Two questions were posed near the beginning of this chapter, and we now have enough information to answer both of them. (a) NIRS *can* be used to analyze the polymer for the stabilizer over the range 0 to 45 units with a SEE of 4.4 units when both vendor and solvent are taken into account. (b) Of the six factors identified, two are significant (i.e., have high t values) and consequently have contributions well above the “noise” level and should not be ignored.

14.7 WHAT ABOUT MORE THAN TWO LEVELS?

For N levels of a given parameter we must use $N - 1$ indicator variables. Assume, for example, that we have samples of a product manufactured at four different sites. The four sites can be identified with three indicator variables, designated $IV1$, $IV2$, and $IV3$ as follows:

Site	Ind. var.		
	$IV1$	$IV2$	$IV3$
A	0	0	0
B	1	0	0
C	0	1	0
D	0	0	1

Note that one site is arbitrarily designated as the “reference” site, and all its indicator variables are given the value zero. For all of the samples at one of the other sites, $IV1$ gets the value 1 and all the other indicator variables get 0. The pattern continues for all sites, with only one “1” appearing for any indicator variable for any given site. Thus, three indicator variables can accommodate four groups of samples.

As soon as the analyst is satisfied that a particular indicator variable is not significant, then it can be dropped from the data input. For example, in the previous listing of four sites, if $IV1$ and $IV3$ were not statistically significant, then it would be clear that site C (identified with $IV2$) is indeed *different from* the three other sites, and samples from site C should be so designated such that the special contribution due to location (whether the influence is positive or negative) can provide the necessary correction.

Here are two more examples: an antibiotic and a paper product.

14.7.1 ANTIBIOTIC

A pharmaceutical company produces a certain antibiotic using four different tanks. From these, 15 liquid samples were drawn and two injections into a liquid-handling flow cell on an NIR instrument were made, providing a total of 30 samples (sufficient for this preliminary evaluation). There was no sample preparation except filtration of the fermentation broth. The data for these samples are shown in the following table:

Tank	Ind. var.					
	1	2	3	Age	N	Conc.
A	0	0	0	4–6	3	82–100
B	1	0	0	5–6	3	77–98
C	0	1	0	0–4	5	1.5–74
D	0	0	1	0–3	4	0–51

Tank designations are arbitrary, age is given in days, and concentrations have been scaled for proprietary reasons. The equation obtained looked similar to this:

$$\begin{aligned} \text{Conc.} = & \sum_{i=1}^3 [F(i) \times \text{Abs}(WL(i))] + F(o) \\ & - 17.8 \times IV(1) \\ & - 30.0 \times IV(2) \\ & - 45.2 \times IV(3) \end{aligned} \quad (14.1)$$

Here's what the equation tells us: Relative to the arbitrary reference tank A, tank B (indicator variable #1) must have 17.8 units subtracted from its uncorrected concentration. Similarly, values from tanks C and D must be *lowered* by 30 and 45 units, respectively. The "error" by ignoring which tank is the source is appreciable.

Additional interpretation is possible when we have more statistics:

$$\begin{array}{lll} \text{Range: 0--100 units (arbitrary)} & \text{Range: } \frac{38.6}{3.34} = 1.6 & F = 616 \\ \text{SD (range)} = 38.6 & & R = .997 \\ \text{SEE} = 3.34 & & \end{array}$$

As a general rule, the "standard deviation of the range" should be at least ten times greater than the SEE. In this example that requirement is met. Also, the F value and the correlation coefficient are sufficiently high to warrant further investigation of NIR as an analytical method for these samples.

If the coefficient associated with an indicator variable is close to the system error (the SEE), then that parameter can probably be ignored. However, if it is more than 2.5 times the SEE, then we do ourselves a favor by including it in the equation. Consider the following data where we calculate this ratio, in other words the contribution as determined by the indicator variable divided by the SEE (the system "noise"):

Tank	IV	Coef.	t Value	Ratio
2	1	17.8	8.6	5.3
3	2	30.0	11.6	9.0
4	3	45.2	15.8	13.5

$$\text{Ratio} = \frac{\text{contribution}}{\text{noise}} \quad \text{where "noise" = SEE} = 3.34$$

The t value alone dictates that the coefficient is significant. The calculated ratio confirms this by showing us the extent to which the contribution exceeds the SEE. (Recall from your first course in analytical chemistry that the limit of sensitivity is typically 1.5 to 2.0 times the system noise. All we're saying here is that values that are several times the system noise are worth keeping.)

14.7.2 PAPER PRODUCT

Fifteen solid samples (two orientations of each) were obtained from two different machines. It was presumed that the machines were performing identically, and the main thrust of the evaluation was to determine the appropriateness of NIR to perform a rapid analysis without any sample preparation (except cutting and/or folding of the paperlike material). While the results were gratifying, the

additional information obtained by employing indicator variables was unexpected and provided an explanation for seemingly erratic analytical results from earlier work. Consider the following:

Machine	IV	N	Concentration	
			A	B
X	1	7	0–100	54–100
Y	0	8	0–100	54–100

Separate equations were generated for each concentration range, leading to an overall better calibration. For the 0 to 100 concentration range, the equation was

$$\text{Conc. (A)} = \sum_{i=1}^3 [F(i) \times \text{Abs}(WL(i))] + F(o) + [12.3 \times IV]$$

For Conc. (B) the factor in the bracket was 48.6. What all this tells us is compiled in the following table:

	Concentration	
	A	B
Range	0–100	54–100
SD (range)	35.6	15.2
SEE	3.2	1.54
Ratio	11.1	9.9
F value	531	460
R	.996	.996

Again, the ratio of SD (range) to SEE is near or above ten, and the other statistics are satisfactory. The following summary suggests that the difference between the two machines is worth noting for the B concentration range; this may or may not be true for the A range:

Range	Coef.	t Value	Noise ratio	Error at top of range (%)
A	12.3	2.5	3.8	12.3
B	48.6	8.0	31.6	48.6

14.8 SUMMARY

Indicator variables can easily be added to a calibration if one has reason to question the influence of one or more controllable factors on the accuracy or precision of an analysis. A factor can be designated “absent” or “present” with assigned values of 0 or 1, or additional indicator variables can be used to designate more than two possibilities. The evaluation of the statistics produced by incorporating indicator variables provides quantitative evidence of the effect that one or more factors may have on a given analytical procedure or process. Thus, this can be a route to

- Increased accuracy or precision
- Simplified sample handling by elimination of irrelevant steps
- Revealing previously unknown interactions in a process or procedure

FURTHER READING

Spectroscopy Europe, 11(2), 35–37 (1997). Available at www.spectroscopyeurope.com/TD_11_2.pdf

15 Qualitative Discriminant Analysis

Howard Mark

CONTENTS

15.1	Introduction	307
15.2	Methods Based on Selected Wavelengths: Use of Non-Euclidean Distance Measures ..	308
15.2.1	Wavelength Selection	316
15.2.2	Normalization.....	317
15.2.3	Applications.....	318
15.2.3.1	Detection of Samples Not Included in Training Set.....	319
15.2.3.2	Warnings of Misclassification	320
15.2.3.3	Transfer of Calibrations.....	321
15.2.3.4	Applications Involving a Single Group	321
15.3	Methods Based on Use of Entire Spectrum	327
15.3.1	Use of Direction Cosines	328
15.3.2	Use of Direct Spectral Matching.....	328
15.3.3	Use of Principal Components	329
15.4	Comparison of Methods	330
	Acknowledgments	331
	References	331

15.1 INTRODUCTION

All qualitative spectroscopic analysis depends in one way or another on comparing spectra of the specimens to be identified with spectra of “knowns” or “standards.” In the course of development of the various spectroscopic methods used for chemical analysis, many methods of making such comparisons have been devised. A small subset of these methods has been applied to near-infrared (NIR) analysis. The methods have been chosen by the various NIR practitioners to match the characteristics of the NIR spectra to which they are applied.

The methods used all have one characteristic in common: They avail themselves of the powerful computer hardware invariably associated with the NIR spectrometers used. They all apply one of several powerful algorithms that allow accurate identifications to be made by distinguishing what are often small absorbance differences. The enormous signal-to-noise ratios that the instruments provide are, of course, indispensable, but even with this, a computer is needed to separate information that cannot always even be detected with the unaided analyst’s eye. Thus, in the NIR region, using different analytical methods is equivalent to applying different algorithms to the data.

Having seen what is the same, let us now consider what is different between the different algorithms used. One way to distinguish among the algorithms is to separate them by the number of wavelengths that are included in the matching process.

Reflecting the history of NIR analysis as having been developed from both monochromator-based and interference filter-based spectrometers, some algorithms match spectra across the entire spectral

region, while others use relatively few individually selected wavelengths from the spectrum, with an auxiliary algorithm that optimizes the selection. We will begin by considering methods that use selected wavelengths.

15.2 METHODS BASED ON SELECTED WAVELENGTHS: USE OF NON-EUCLIDEAN DISTANCE MEASURES

The most common method of classifying an unknown material using optical spectroscopy is to visually examine the spectrum. The chemist inspecting a spectrum for the purpose of determining the nature of the sample giving rise to the observed bands normally concentrates his attention on the regions of the spectrum showing absorbance peaks, and classifies or identifies the sample by matching the location and strength of absorbance peaks to those of known substances. It is possible to generalize this procedure by noting that, if the absorbance is measured with sufficient accuracy, then any wavelength where there are absorbance differences between the substances to be distinguished can serve to classify them.

For illustration, we consider the three spectra in Figure 15.1, which represent the spectra of hard wheat, soft wheat, and soy meal. Figure 15.1a shows the NIR reflectance spectra of these materials over the wavelength region 1100 to 2500 nm. Figure 15.1b shows one section of the same spectra, expanded so as to emphasize the portion of the spectrum on which we wish to concentrate our attention. Figure 15.1b also indicates the approximate high and low extremes of several readings of each sample. The spread between the spectra of each material is characteristic of reflectance spectra of powders. When repacked, different grains of the powder are near the surface, and the differences in orientation, particle size, etc. of the surface grains gives rise to differences in reflectance when different packs of the same material are measured. Thus, in the case of solid powders measured by reflectance, the generation of discriminating criteria is complicated by the need to distinguish the different materials in the face of this extraneous source of variation in the measured spectra.

In this case it is easy, almost trivial, to distinguish these materials by eye: The presence of high concentrations of starch in wheat causes the measured absorbance at 2100 nm to always exceed the absorbance at 2180 nm, while the absence of starch in soy meal causes the reverse relation to exist in that material. Due to the nature of the different types of wheat, the kernels of hard and soft wheat grind differently, so that there always exists a large particle size difference between the ground materials, hard wheat always having the larger particle size when the normal grinding is carried out (Udy Cyclotec grinder fitted with a 1-mm screen). This difference in particle size is converted to a corresponding difference in measured absorbance value upon measurement of the NIR spectra of the two materials. As Figure 15.1b shows, this difference is larger than the difference due to repack of either material. Thus the two wheat types can be distinguished by their overall level; this can be done at any arbitrary wavelength, or, in particular, it can be done at the same two wavelengths that, as we have seen, serve to distinguish wheat from soy meal.

So much for our visual observations of the spectra. Following the earlier development of this topic [1,2], we want to generate a method of converting these visual observations to something that can be fed into a computer, so that the spectral inspections and classification tasks can be done automatically. In order to do this we note the following: Each spectrum is really “only” a plot of the absorbance of the samples at each wavelength. In the innards of the computer these same spectra are represented by the numerical values corresponding to the absorbance of the materials at each wavelength. We have already noted that we can perform the task of distinguishing these three materials by using only two wavelengths: 2100 and 2180 nm. Therefore it should be possible to dispense with all the data except the absorbances of the specimens at those two wavelengths and still be able to classify the samples.

Table 15.1 is a partial listing of the data from the wheat and soy meal. From the data in Table 15.1 we note that each spectrum has an absorbance value at the two wavelengths we will use to perform the classification, 2100 and 2180 nm, that conforms to the conditions we noted from the spectra.

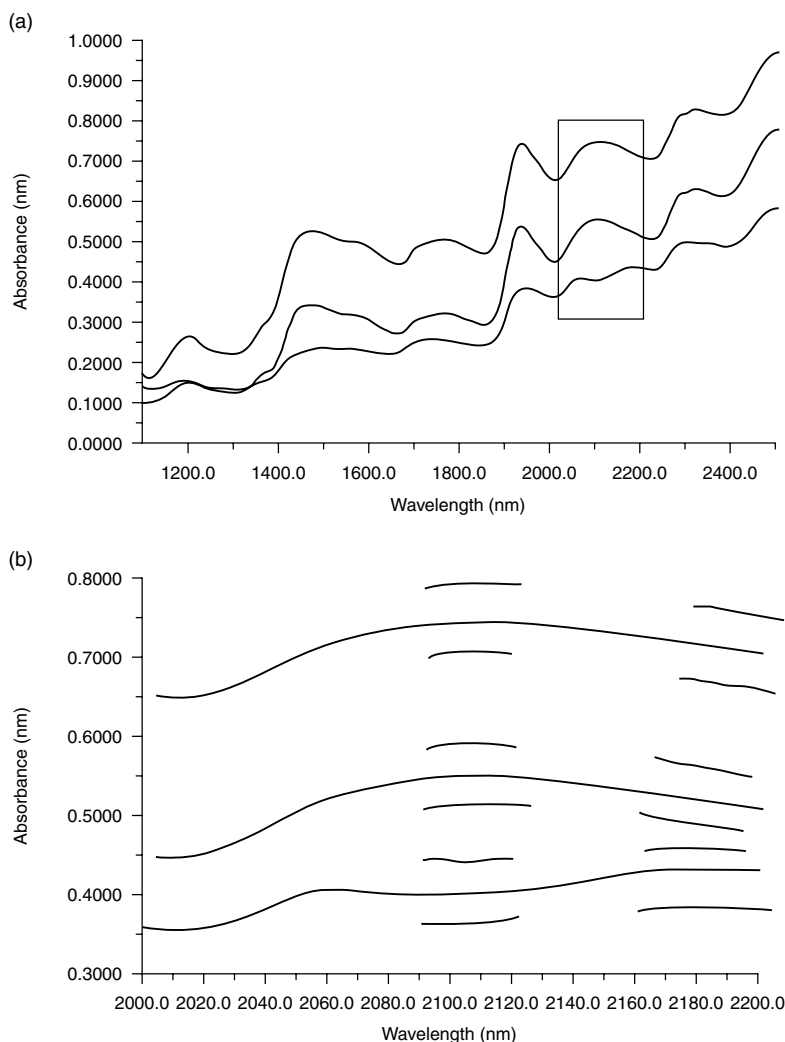


FIGURE 15.1 Spectra of hard wheat, soft wheat, and soy meal. (a) Spectra over the full range of the instrument, 1100 to 2500 nm. (b) Boxed portion of part (a), expanded.

Now, every scientist knows what can be done with tabular data such as those presented in Table 15.1: we can make a plot. By plotting the absorbance readings of several samples of each of the three materials at 2100 nm vs. the corresponding readings at 2180 nm, we obtain the results shown in Figure 15.2a. The three materials are represented by three groups of points. The groups are well separated; this indicates that the spectra of the various materials are sufficiently different at these two wavelengths that, as we have seen in the previous spectral plot, these two wavelengths alone can be used to characterize the three materials. All we would need to do to analyze unknowns would be to compare their spectra with those of these three known materials by plotting the point corresponding to the absorbances of the unknowns at these same two wavelengths. If the point corresponding to the unknown fell among those for any of the knowns, we would classify the unknown as being the same as the known material within whose data the point fell.

Clearly, more substances could be included in this set of materials to be distinguished as long as all were different at these two wavelengths; in such a case they would be represented as points in different parts of the plane containing the data.

TABLE 15.1
Partial Listing of Data Corresponding to Typical Absorbance
Readings of Hard and Soft Wheats and Soy Meal at the Two
Wavelengths Used for Discrimination^a

Hard red spring wheat		White club wheat (soft)		Soy meal	
2100 nm	2180 nm	2100 nm	2180 nm	2100 nm	2180 nm
0.50482	0.46430	0.42552	0.37907	0.51368	0.54222
0.49690	0.45664	0.43536	0.38604	0.52308	0.55082
0.48190	0.44505	0.43257	0.38206	0.50879	0.54411
0.50106	0.46028	0.43321	0.38555	0.53371	0.56455
0.48669	0.44647	0.43180	0.38079	0.50434	0.53654
0.50616	0.46568	0.42996	0.38134	0.48316	0.51229
0.50186	0.46176	0.41947	0.37417	0.53172	0.56585
0.49157	0.44906	0.42050	0.37332	0.46237	0.49642
0.51388	0.47172	0.42082	0.37219	0.55124	0.58448
0.52275	0.47966	0.42517	0.37823	0.51540	0.54384

^a Extracted from the data used to generate Figure 15.2.

In general, of course, we could not expect to classify an arbitrarily large number of different materials using only two wavelengths. Rather, we would need many wavelengths, corresponding to a large number of dimensions. For three wavelengths, we could conceivably draw a picture similar to that in Figure 15.3. We would require three axes, one for each wavelength. Each axis would require one dimension to contain it; therefore a three-wavelength discriminating function requires three dimensions to represent it. We also require of the data one more characteristic that we mention for the sake of completeness: the data for each group should have a multivariate normal distribution. In this case the cigar-shaped regions of space shown in Figure 15.3 are the regions containing data points corresponding to the absorbances at three wavelengths. The absorbance at any given wavelength determines the position of the point for that specimen parallel to the corresponding wavelength axis. Data points for similar materials form clusters in what is now a three-dimensional space — one dimension for each wavelength used for the identification.

Use of more wavelengths would require correspondingly higher dimensions. For large numbers of wavelengths, which require correspondingly large numbers of dimensions, we must abandon the visual approach and create a mathematical method for locating data in multidimensional space. Even for three dimensions, while we can draw schematic diagrams to illustrate the concepts, it is not really practical to actually try to plot real data and draw conclusions from such a three-dimensional plot. A computer can handle any number of dimensions with equal ease; so in all cases beyond an ordinary two-dimensional case, it is much more convenient to devise such a mathematical method so that we can discriminate between different materials using as many wavelengths as necessary to distinguish them. The two- and three-dimensional cases, however, serve to bridge the gap between our visualization of what happens and the mathematical representation of it.

We will generate this mathematical description for the two-dimensional case, thus relating it to the visual description. Consider Figure 15.2a and note that each group is located in a different part of the space of interest. We will classify samples by defining the locations of the various groups in this space, and assign a sample to a group if it is “near” that group. This is simply a matter of quantifying the visual approach mentioned previously, where we concluded that we would identify an unknown if it fell “within” the group boundaries; previously we did not specify how we would determine if a new data point was within the boundaries or not.

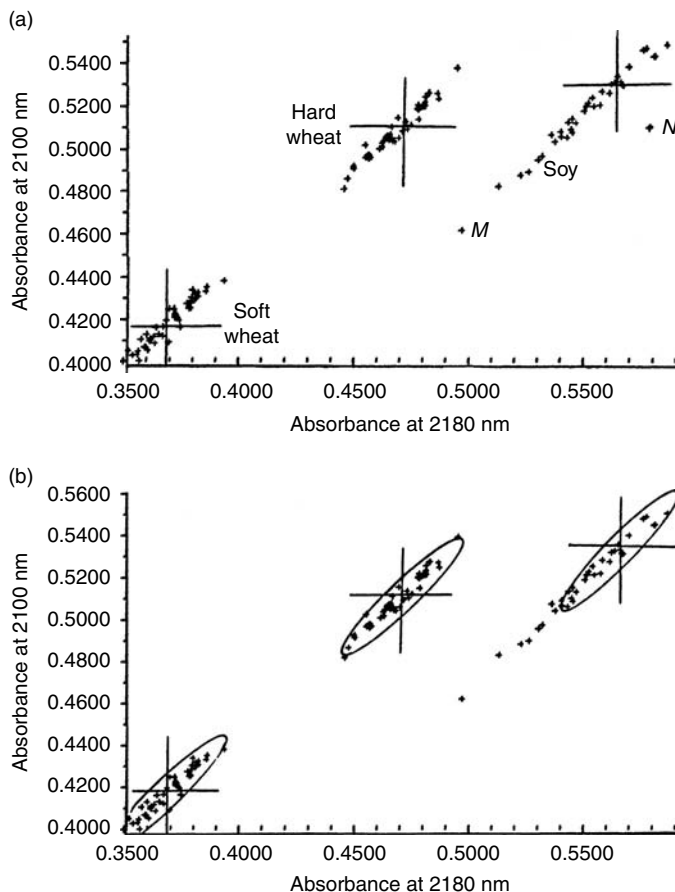


FIGURE 15.2 Plots of the absorbances of the three materials at the two wavelengths of interest. Since similar materials have similar relations between the wavelengths, they tend to cluster together, forming groups in the two-dimensional space. In part (a) the means of each group at each wavelength is indicated; where they cross is the “location” of the group.

The problem thus breaks down into two parts: locating the groups, and defining a criterion for determining whether a sample is near a given group in the space.

We define the position of a group in multidimensional space as the point corresponding to the mean value at each measured wavelength. This is illustrated in Figure 15.2a, where the mean at each wavelength for each group is indicated; the point where the lines join at the center of the group is the “location” of the group in space. The list of numbers describing the mean absorbance values of any given material (or the corresponding group of data points) at the several wavelengths being used is a mathematical vector called the *group mean* of that group. The set of vectors describing all the materials to be distinguished is called the *group mean matrix*.

The determination of distance is a bit more complex. Euclidean distances are not optimum, for the following reason: If we consider the three groups in Figure 15.2, we note that each group is somewhat elongated, with the direction of elongation lying along an approximate 45° line compared to the axes of the plot. If we consider points M and N in Figure 15.2a, it is clear that point M is likely to be a member of group B, while point N is not, because point M lies within the cluster of points that cumulatively define group B while point N does not. This is despite the fact the point N lies closer to the central location of the group than point M does.

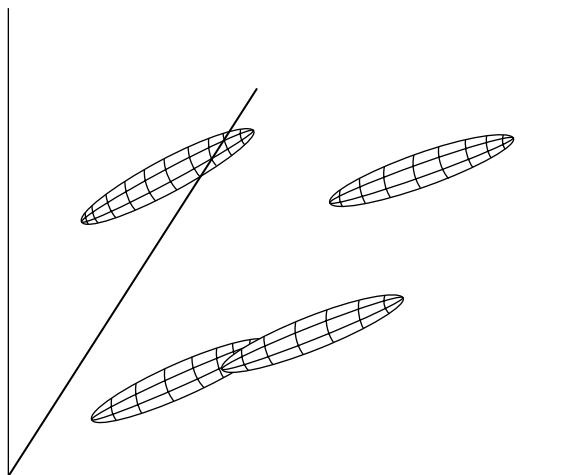


FIGURE 15.3 Ellipsoids are the three- and higher dimensional surfaces that define the size, shape, and orientation of the data arrays in the corresponding multidimensional spaces. These surfaces are the equivalents of the ellipses that in Figure 15.2 do the same job in two dimensions.

To overcome this difficulty, we would like to define a distance measure D in such a way that the equivalent Euclidean distance is large in those directions in which the group is elongated. A method of accomplishing this was introduced by Mahalanobis [3], and the quantity D , defining the “unit distance vector” in multidimensional space, is called the Mahalanobis distance.

The Mahalanobis distance can be described by an ellipse (or ellipsoid, in more than two dimensions) that circumscribes the data, as shown in Figure 15.2b. A very readable development of the mathematics is presented by Gnanadesikan [4]. The distance D , from a point X to the center of a group i , is described by the matrix equation:

$$D^2 = (X - \bar{X}_i)' M (X - \bar{X}_i) \quad (15.1)$$

where D is the distance, X is the multidimensional vector describing the location of point x , \bar{X}_i is the multidimensional vector describing the location of the group mean of the i th group, $(X - \bar{X}_i)'$ is the transpose of the vector $(X - \bar{X}_i)$, and M is a matrix determining the distance measures of the multidimensional space involved.

The relationship of this equation to the previous discussion is apparent: The various sets of \bar{X}_i represent the group means of the different materials to be distinguished. The distance measures described are defined by the matrix M . It is convenient to consider the distance from the center of the group to the circumscribed ellipse as representing 1 SD of the data.

We still need to define the boundary of a group. An initial “ballpark” value for this criterion was set at three times Mahalanobis distance, but recent work on the problem has shown that a more accurate value can be determined by considering the number of wavelengths and the number of spectra in the training set [5]. In this work it was shown that the upper control limit (UCL, i.e., the boundary of the group) should be between 2.18 and 6.24 times Mahalanobis distance for common combinations of the number of wavelengths and number of spectra in the training set. Table 15.2 lists the appropriate boundary values for various combinations of the numbers of wavelength and spectra contributing to each group.

There are three possible scenarios that can occur when unknowns are read: the data from the unknown sample can fall within the boundaries of zero, one, or more than one group. If the data from the unknown do not fall within any group boundaries, then the unknown represents a sample that was not included in the training set. If the unknown falls within the boundaries of one group

TABLE 15.2
Values for the UCL to Use for Defining the
Boundary of a Group

No. wavelengths	α level	No. samples in training set			
		25	50	75	100
2	.01	3.42	3.22	3.15	3.12
	.05	2.66	2.55	2.52	2.50
	.10	2.30	2.22	2.19	2.18
3	.01	3.94	3.63	3.54	3.49
	.05	3.14	3.96	2.90	2.87
	.10	2.76	2.62	2.58	2.56
5	.01	4.90	4.33	4.17	4.09
	.05	4.00	3.63	3.52	3.47
	.10	3.57	3.28	3.19	3.15
7	.01	5.89	4.95	4.71	4.60
	.05	4.84	4.21	4.04	3.96
	.10	4.35	3.84	3.71	3.64
9	.01	6.98	5.54	5.20	5.05
	.05	5.75	4.76	4.51	4.41
	.10	5.18	4.37	4.17	4.08
10	.01	7.59	5.83	5.44	5.26
	.05	6.24	5.03	4.74	4.61
	.10	5.63	4.63	4.39	4.28

Source: Reproduced from R. G. Whitfield, M. E. Gerger, and R. L. Sharp, *Appl. Spectrosc.*, 41: 1204–1213 (1987). With permission.

only, then the unknown sample is identified, in other words, it can be assigned unambiguously to that group. If the unknown falls within the boundaries of more than one group, there is an ambiguity. This is indicative of insufficient separation of the different groups, in other words, the spectra are too similar (at least at the wavelengths chosen) to classify the samples; we will discuss this case at length further on. However, in this case we are forewarned of the possibility of misclassification because the groups involved will have their group means within a distance of less than twice the boundary size, making the boundaries of the groups overlap.

Gnanadesikan [4] describes three variant ways to construct the circumscribing ellipses, calling them M1, M2, and M3. These are illustrated in Figure 15.4 (for the two-dimensional case). Figure 15.4a illustrates M1, a unit matrix. The ellipsoid reduces to a sphere, and the use of M1 reduces the calculation of distance D to the calculation of Euclidean distances.

Matrix M2 uses the inverse of the variance-covariance matrix; the elements of the variance-covariance matrix are defined by the calculation:

$$\text{Element } V_{ijk} = \sum_{l=1}^{n_k} (X_{lik} - \bar{X}_{ik})(X_{ljk} - \bar{X}_{jk}) \quad (15.2)$$

where V_{ijk} represents value of i, j th element of the variance-covariance matrix corresponding to the k th group, X_{lik} represents the value of the data from the l th specimen of the k th group at the i th wavelength, and \bar{X}_{ik} represents the mean value of the data for the k th group at the i th wavelength.

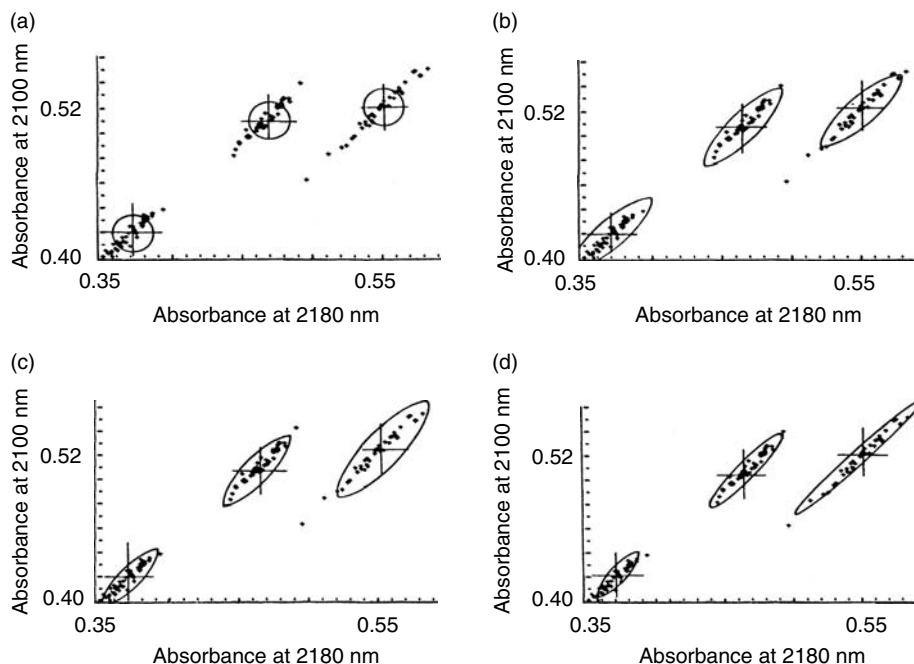


FIGURE 15.4 Variations on the theme — the bounding ellipsoids can be fitted to the data in various ways. (a) Circles: space is Euclidean. (b) Same ellipse for each data group: defines Mahalanobis distance; all groups are assumed to have same size, shape, orientation. (c) Ellipses have same shape and orientation; size is fitted to RMS group size. (d) Each ellipse is fitted individually to the corresponding group, which may each have different sizes, shapes, and orientations.

The summation is taken over all the samples belonging to a particular group. Use of M2 implies fitting a separate ellipsoid to the data for each material to be distinguished. This is demonstrated in Figure 15.4d.

Matrix M3 of Gnanadesikan describes the inverse of the matrix formed by pooling the within-group covariance matrices of all groups. This is an approach that has been implemented in currently available commercial instruments. Use of M3 defines a common metric for all groups in the dataset, indeed for the entire multidimensional space; this is shown in Figure 15.4b. The defining equation is similar to Equation (15.2), but there are two differences: First, there is only one inverse variance–covariance matrix that applies to the entire dataset, and the entire space. For this reason, the same “rubber ruler,” which stretches or shrinks depending on the direction of measurement, applies to the entire space. Second, the elements of this matrix, before inversion, are the summations of the corresponding elements of the individual variance-covariance matrices defined by

$$\text{Element } V_{ij} = \sum_{k=1}^p \sum_{i=1}^{n_k} (X_{lik} - \bar{X}_{ik})(X_{ljk} - \bar{X}_{jk}) \quad (15.3)$$

where the summations are now taken over all the p different groups.

Typical values for a group mean matrix and for a pooled inverse variance–covariance matrix are illustrated in Table 15.3, in which the values have been computed from the wheat and soy data of Figure 15.1 and Figure 15.2, and which is partially listed in Table 15.1.

The statistical theory of discriminant analysis also defines a linear discriminant function very similarly to Mahalanobis distance. These functions have characteristics that are of interest to NIR spectroscopists: the linear discriminant functions are similar in form to regression equations, so that,

TABLE 15.3
Values of the Group Mean Matrix, Inverse Pooled Variance–Covariance Matrix, and RMS Group Size for the Data of Table 15.1^a

	Group mean		RMS group size
	2100 nm	2180 nm	
Hard wheat	0.51008	0.46790	1.24557
Soft wheat	0.41894	0.37029	1.32280
Soy meal	0.51898	0.54992	1.70579
Inverse pooled variance–covariance matrix			
	Elements corresponding to 2100 nm	Elements corresponding to 2180 nm	
Elements corresponding to 2100 nm	114801.9385	–118775.7229	
Elements corresponding to 2180 nm	–118775.7229	128155.7967	

^a Also illustrated in Figure 15.1 and Figure 15.2.

in principal, at least, they could be entered into the current types of instrumentation the same way that standard quantitative calibrations are.

However, the current implementations of the discriminant analysis approach to qualitative analysis via NIR spectroscopy use Mahalanobis distances rather than linear discriminant functions because of their other characteristic: Linear discriminant functions do not permit as straightforward a means of detecting outliers and samples not in the training set as Mahalanobis distances do, in addition to other advantages that will be considered further on.

The main disadvantage of using Mahalanobis distance for actual discrimination of unknowns (compared to using linear discriminant functions) is the amount of computation required. As shown in Equation (15.1), the calculation of a Mahalanobis distance requires two matrix multiplications. This is considerably more computation than use of a linear discriminant function requires, because once the linear discriminant functions have been determined, classifying unknowns using linear discriminant functions requires no more computation than a similar regression equation. However, the continuing decline in the cost of ever-increasing amounts of computer power makes this consideration almost trivial.

The use of these concepts in practice is straightforward. Data at the wavelengths available in an instrument are collected for several samples representing the various materials to be distinguished; these constitute the training set. The data at each wavelength are averaged separately for each material, forming the group mean matrix. The pooled covariance matrix is formed by creating the covariance matrix for each group as described in Equation (15.2), adding the corresponding terms of each individual covariance matrix, and dividing by $n - m$ (where n is the total number of samples and m is the number of groups). This matrix can now be inverted by any of several standard methods. The group mean matrix and inverse pooled covariance matrix in conjunction constitute what can be thought of as the calibration “equation,” even though they do not represent a single actual equation.

To classify an unknown sample, the Mahalanobis distance D from the unknown to each of the materials in the training set is calculated from Equation (15.1). The sample is assigned to the group to which it is closest. In the usual case, a given sample will lie within the boundaries of the group to which it belongs, and far away from any other group. There are two noteworthy exceptions to this. The first exception is the case where a given unknown is outside the boundaries of any group.

Assuming that the discrepancy is not due to statistical variability (which can be overcome by using a boundary corresponding to a lower α level from Table 15.2) or a change in the measurement technique, then this is an indication that the unknown sample is of a type not represented in the calibration teaching set.

The other exception is the case where a given sample is close to more than one of the calibration samples. This can only occur if the calibration samples themselves have very similar spectra and thus are close together in multidimensional space. This can be detected during the calibration. As discussed previously, groups closer than twice the boundary distance overlap, potentially causing such misclassification. This situation can be detected during calibration, and steps taken (either mathematical, such as using more and/or different wavelengths; or physical steps, such as different sample preparation methods).

15.2.1 WAVELENGTH SELECTION

When only a few materials are to be distinguished, the differences between the spectra can be ascertained visually and the wavelengths to use for the discrimination selected manually. If some of the spectra are very similar, or if there are so many different materials involved that visual inspection becomes confusing, then such an a priori selection would not be possible. In the absence of a prior knowledge of which wavelengths are suitable for performing the desired discriminations, a method of selecting the optimum set of wavelengths is needed.

A method that has been used in the statistical literature has been to compute the value of

$$\sum_{k=1}^p (\bar{X}_k - \bar{\bar{X}})' M (\bar{X}_k - \bar{\bar{X}}) \quad (15.4)$$

where \bar{X}_k is the vector representing the group mean of the k th group, $\bar{\bar{X}}$ is the vector representing the grand mean of all the groups in the dataset, and M represents whichever matrix defines the distance metric for the multidimensional space. The result of this computation is effectively a comparison of the between-group distances to the within-group distances of each specimen's data from its proper group mean. The variables (i.e., the wavelengths or their analogs) are chosen so that the between-group distances are maximized. This procedure is useful in some of the discrimination problems that statisticians deal with, but has one major drawback for spectroscopic analysis: the between-group distances calculated according to Equation (15.4) depend most heavily on those groups that are far away from the others, in other words, those whose spectra are most different. However, materials with greatly differing spectra are easy to distinguish, almost regardless of the wavelengths used. What is needed in a practical discrimination situation is a method of maximizing the distances between groups that are close together in multidimensional space, in other words, a criterion that emphasizes the differences between those materials that have similar spectra.

The statisticians' solution to this problem has been to compute all the intergroup distances and select the variables that maximize the distance between the closest pair of groups.

The method of wavelength selection that has been devised for implementation with the current algorithms in use in NIR is similar to the second method described above: the distances D_{ij} between all pairs of groups i and j are computed, then the sum of the inverse squared distance, in other words, $(1/D_{ij})^2$ is formed. The groups that are closest together will contribute most heavily to this sum; thus selecting those wavelengths that cause this sum to be smallest results in the selection of the wavelengths that best separate the closest groups, in other words, best distinguish the most similar spectra. The groups that are far apart (i.e., dissimilar) are no problem to distinguish among. This approach has an advantage over the simpler one in that this technique will optimize among all groups that are comparably closely spaced rather than concentrating on only the single closest pair.

15.2.2 NORMALIZATION

The use of the M3 matrix, in other words, the inverse pooled covariance matrix, as the basis of calculating Mahalanobis distances in multidimensional space implies that all the groups have essentially the same size, shape, and orientation, so that only one specifying matrix is needed to describe them all. In NIR analytical technology, as we have seen, the nature of the reflectance phenomenon is such that the data at all wavelengths have a tendency to increase or decrease together. As demonstrated in Figure 15.2, this shows up as the tendency for the data points to lie along parallel lines. In three dimensions, the phenomenon would appear as a tendency for the data of the various groups to occupy parallel “needles” or cigar-shaped regions in space (shown in Figure 15.3), and so forth. However, empirically we find that materials with large values of $\log(1/R)$ have larger variations in the data than samples with small values of $\log(1/R)$. This suggests that, while different groups tend to have the same shape and orientation, they can differ in size. This could lead to false conclusions when analyzing unknowns, since pooling the data from large and small groups gives an incorrect estimate of the sizes of the individual groups. This could then lead to false rejection of samples that are actually part of a large group, and false assignment of unknown samples to a small group when in fact the sample lies beyond the boundary of that group.

To avoid this problem, the Mahalanobis distance can be normalized according to the sizes of the various groups that are in the training set. Having calculated M3, we can compute the root mean square (RMS) size of each group by first calculating the distance D_i of each sample in the training set from the group mean of its assigned group; then, for each group, calculate:

$$\text{RMS group size} = \frac{\sum D_i^2}{N - 1} \quad (15.5)$$

Dividing the Mahalanobis distance from a sample to each group mean by the RMS group size for the corresponding group will normalize the distances so that more accurate representations of the distance to the various group means can be calculated, decreasing the possibility of error.

For the wheat and soy data used for the previous illustrations, the RMS group sizes are also presented in Table 15.3. Comparison of these numbers with the plotted groups in Figure 15.2 will help relate the mathematical concepts to the visual appearance of the data.

Computing normalized distances in this manner provides a result that is, in a sense, intermediate between the results obtained through use of the M2 and M3 matrices of Gnanadesikan. Whereas M2 assumes a different shape, size, and orientation for each group, and M3 assumes the same shape, size, and orientation for all groups, the use of normalized distances assumes the same shape and orientation, but different sizes for each group in the calibration. Figure 15.4c illustrates the relation of normalized distances to the other multidimensional measures; these relationships are also summarized in Table 15.4.

Note that the various matrices defining the distance measures in multidimensional space are in order of increasing improvement of the fit of the defining ellipsoid to the data groups, starting from the completely arbitrary hyperspheres, to fitting to the pooled group shapes, to fitting the groups sizes, to fitting each group individually. This increasingly improved fit of the ellipsoids to the data provides added accuracy of classification, but comes with concomitant side effects, as shown in the various columns of Table 15.4. The side effects include a requirement for increasingly large numbers of calibration samples, and a change in the nature of the space to becoming non-Euclidean and nonmetric spaces.

Use of the unit matrix, matrix M1 using the nomenclature of Gnanadesikan, is the only space that is Euclidean. The use of circles (or spheres, or hyperspheres, depending on the dimensionality) gives such spaces the characteristic of having the same measuring “yardstick” in all directions. In all other cases, the use of an ellipse (or ellipsoid) means that the yardstick stretches or contracts depending on the orientation. However, in a given direction, the measure of distance is uniquely

TABLE 15.4
Comparison of the Various Multidimensional Distance Metrics

Symbol	Defining matrix	Nature of ellipsoid	Number of samples required for calibration	Nature of space
M1	Unit matrix	Circles (hyperspheres) implies all groups are spherical	Five per group, to define group mean	Euclidean metric ($D_{ij} = D_{ji}$)
M3	Pooled inverse variance–covariance matrix	Same ellipsoid for all groups; implies all groups have same size, shape, orientation	Five to ten per group; pooling several groups provides good estimate of ellipsoid	Non-Euclidean metric
Normalized M3	Pooled inverse variance–covariance matrix, modified by RMS group size	Modified ellipsoid; implies all groups have same shape and orientation, different sizes	10 to 15 per group or enough to ensure good measure of one groups “standard deviation”	Non-Euclidean nonmetric ($D_{ij} \neq D_{ji}$)
M2	Individual inverse variance–covariance matrix	Ellipsoid fitted to each group implies each group has different size, shape, and orientation	25 per group, or enough to define each group’s shape and orientation	Non-Euclidean, nonmetric

specified, and such spaces are called *metric spaces*. Use of matrices M2, M3, or normalized group sizes all imply non-Euclidean spaces.

Use of the unit matrix, M1, or the pooled inverse variance–covariance matrix, M3, further implies that distances measured in opposite directions are the same, in other words, $D_{ij} = D_{ji}$. This does not hold true for distances measured in spaces where the group sizes are normalized, or where the inverse variance–covariance matrices are computed separately for each group. A measurement from group i to group j uses the distance measurement appropriate for group i , while that from group j to group i uses the distance measurement appropriate for group j . Spaces such as these are called nonmetric spaces because there is no unique specification of the distance measure, even in a given direction, and $D_{ij} \neq D_{ji}$.

The characteristic of a space as being nonmetric does not preclude the use of such a space for performing classification. When calculating distances in a nonmetric space, the distance from a sample to a given group is calculated using the distance measure appropriate to that group; this provides an unambiguous determination of that distance. Put into different terms, if we go back to thinking of the multidimensional distances as standard deviations, this is equivalent to calculating how many standard deviations away from each group a given sample is. And the fact that each group has a different standard deviation is of no consequence: we simply measure from each group in terms of its own standard deviation.

15.2.3 APPLICATIONS

The most obvious and straightforward use of these concepts is the application to qualitative analysis. Having generated a set of group means and an inverse pooled variance–covariance matrix for a given set of materials, it then becomes possible to determine whether unknown materials can be classified as one or another of the knowns, by calculating the distance of the unknown to each of the known

groups, and assigning it to one group (or, to use equivalent chemical terminology, to *identify* it as one of the knowns by determining that the spectra match) if it falls within the boundaries of that group.

In discussing the concepts used to set up the multidimensional classification scheme, we have been using the terminology and nomenclature of statistics, with occasional references to the equivalent terminology from chemistry, chemometrics, and/or spectroscopy. It is pertinent at this point to indicate how the previous discussion relates to these more familiar subjects.

In terms of chemometrics, the concepts presented here fall into the category of *supervised* classification methods, because it must be known beforehand which group each specimen falls into. Unsupervised training methods also exist, in which samples are grouped based solely on the characteristics of the data. In order to use the methods presented here, the additional information, in other words, which samples belong to which group, must be known beforehand. It is conceivable, however, that that information could be derived from an unsupervised classification scheme and then applied to the supervised scheme.

In the nomenclature of chemistry and/or spectroscopy, these methods fall into the categories of library searching and spectral matching. It must be recognized, however, that the library may be a rather specialized one, and the concept of a spectrum comes close to breaking down; absorbances at only a few disparate wavelengths is only marginally a spectrum, while the absorbance at only one wavelength may perhaps be considered a “spectrum” only by courtesy.

15.2.3.1 Detection of Samples Not Included in Training Set

As was discussed, data from a material not in the training set, which has different spectral characteristics from any of the included materials, will fall outside the boundaries of all of the groups. Figure 15.5 illustrates such a case, where we have added a synthetic sample represented by artificial data, that falls outside all the group boundaries. The distances between unknown sample to each of the known groups are listed in Table 15.5.

From Table 15.5 it is clear that the new sample is not part of any group, indicating that such a sample has not been seen before. A caution is in order here though. It is possible that the sample is not

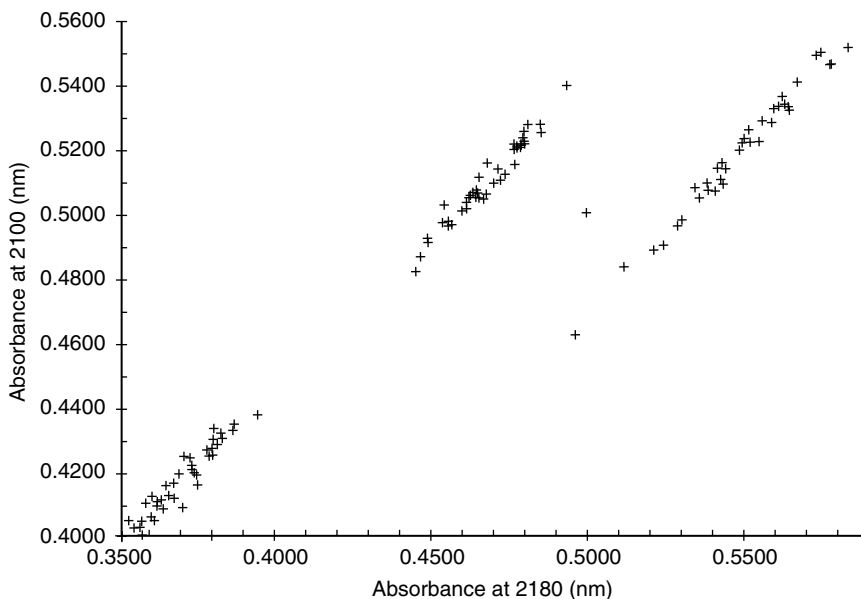


FIGURE 15.5 The data point outside any of the defined groups has a large Mahalanobis distance from all of them, and would be classified as not belonging to any of the groups in the training set.

TABLE 15.5
Distance between the Artificially Introduced
Data Shown in Figure 15.5, Outside Any of the
Known Groups, and Each Group

Group	Mahalanobis distance	Normalized distance
Hard wheat	14.85	11.92
Soft wheat	20.32	15.36
Soy meal	11.65	6.83

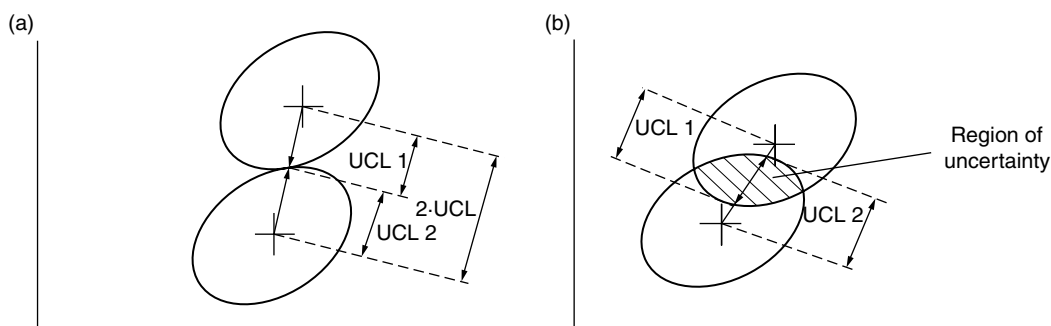


FIGURE 15.6 (a) When two groups are separated by exactly twice the UCL, the group boundaries touch and are on the verge of overlapping. (b) When the two groups are closer than twice the UCL, the boundaries overlap; any data falling in the shaded region cannot be unambiguously assigned to either group.

consistent with the known data groups due not to its being an unknown material, but because of an anomalous effect, such as a shift in the operating point of the instrument or, perhaps, if the sample is a solid, a worn grinder or improper packing of the sample into the cup. It may also indicate a sample that is contaminated, rather than one that is completely different from what it is supposed to be.

15.2.3.2 Warnings of Misclassification

Table 15.2 lists the UCLs for Mahalanobis distances, in other words, the distance from the group mean defining the outer boundary of each group. If we perform the thought experiment of imagining two groups approaching each other, the regions of multidimensional space defining the two groups will be isolated from each other as long as the two group boundaries do not overlap. As the groups get closer and closer, eventually the two boundaries will touch each other. As shown in Figure 15.6a, this condition occurs when the distance between the two groups is exactly twice the UCL.

If the two groups get still closer, then the regions of space defined by the group boundaries overlap. A datum from a sample falling in this region cannot be assigned unambiguously to either group, because it falls in the domain of both groups. This situation occurs when the distance between the two groups is less than twice the UCL, and is illustrated in Figure 15.6b.

This condition can be detected by computing intergroup distances for all pairs of materials during the calibration step. What to do about it is not as readily discerned. If the groups are “too close,” it means that the spectra of the materials are too nearly similar, at least at the wavelengths used, to tell the difference between them. If other wavelengths are available that show larger differences, then these should be included; that will improve matters. If, on the other hand, there are no other wavelengths available with larger differences than the ones already in use, then things become more difficult.

It is almost unheard of in NIR analysis to admit defeat; someone is always holding out the carrot of a “magic” wavelength set, or a “magic” data treatment that will resolve all problems. The fact of the matter is that if the spectra are the same, nothing you do to the data will make them be different. As we will see in Section 15.4.1, however, including wavelengths that do not have discriminatory power will make things worse rather than better.

Is there any hope at all, then, in such cases? Yes, but it becomes necessary to turn to other areas of science for the answers. At this point no specific recommendations are available because the NIR community has not yet started to include other types of expertise in solving NIR problems. In general, however, there are some areas that can be pointed to.

Making use of chemical derivatization (as distinguished from mathematical derivatives: $dA/d\lambda$) has been used for many years in distinguishing chemicals that are outwardly similar. Causing a chemical reaction to occur that gives rise to very different products from similar starting materials is an old and hallowed methodology. We generalize this procedure by noting that finding a way to treat the *samples* to enhance the differences between them will work, and more surely than trying a whole bunch of empirical data treatments. The treatment may be chemical or physical; the nature of what is needed will depend on the materials.

This, of course, is much more work than simply putting samples into an instrument or data into a computer. However, in difficult cases, the instrument and/or the computer may need a little help from the analyst. This, of course, is the point: such manipulations need only be done in the difficult cases. If, out of a large set of materials, two or three cannot be distinguished from each other, then the proper course of action is to define them as single group; then only if a sample is identified by the classification algorithm as belonging to that group is it necessary to take further action, in order to make a final determination as to the identity of the material.

15.2.3.3 Transfer of Calibrations

Having generated a discriminant calibration, it is often useful to be able to utilize that calibration on several different instruments. Because the differences between different instruments result from unavoidable manufacturing variations, the effect on the readings is to cause changes in the value of the readings taken at any given wavelength. The size and shape of the groups in multidimensional space is due mainly to the differences in reflectivity of the sample as different repacks, samples of different compositions, etc., are measured.

To a first approximation, then, we can separate the effects due to instrument and sample; instrumental differences affect mainly the group mean matrix, and within-group sample differences affect mainly the inverse pooled covariance matrix. Thus, in order to use the Mahalanobis distance-based discriminant function on a different instrument than the one it was generated on, it is necessary to change only the group mean matrix to accommodate the readings from the new instrument.

This adjustment can be performed on fewer samples than were needed to perform the original calibration. In favorable cases only one sample of each material is needed. The “favorable case” in this instance is that the sample of each material gives a reading that is a good approximation to the mean value of the readings, at each wavelength included in the calibration. Otherwise, several samples or, in the case of powdered solids, at least several repacks should be measured in order to obtain a good estimate of the value of the new group mean vector for each material.

The group mean matrix is then simply replaced with the new values of the mean absorbance for each material to be distinguished.

15.2.3.4 Applications Involving a Single Group

As described, a major use of the concepts presented in this section is the identification, or discrimination, of one set of materials from the others. This implies the existence of two or more materials; presumably each material will form a distinct grouping of data in multidimensional space. This

presumption will be correct when the spectra of the different materials differ from each other, and there will be as many groups in space as there are different materials to be distinguished (with exceptions as described in Section 15.2.3.2).

On the other hand, it is also possible to perform the calculations that define the circumscribing ellipses using data that correspond to only one group; this will result in a group mean vector and an inverse covariance matrix that define the size, shape, orientation, and position of that single group. Such a detailed description can also be put to a number of uses.

Application to Quality Control and Wavelength Selection for Discrimination from a Single Group. One obvious extension of the use of Mahalanobis distances to classify materials is its application to using NIR to distinguish “good” (or acceptable) product from “bad” (or defective) product for whatever criterion this distinction is based on. As long as the bad product has spectral features that distinguish it from the acceptable product, then qualitative NIR analysis can be used as a quality control tool. If desired, the different ways that any given product can become defective can be used to set up the different data groups, just as if they were actually different materials, because, in fact, they are.

However, more to the point in this section is the fact that such a complicated scheme is not necessary; it suffices to simply define the multidimensional boundaries of the good product; then any sample of future production that falls outside those boundaries is defective (or at least suspect).

Only one new difficulty appears: how to choose the wavelengths that will allow this discrimination task to be performed.

The method of selecting wavelengths presented, for the case of distinguishing among many groups, works satisfactorily for the task for which it was generated, but less well for other tasks. In order to apply this criterion, *all* intergroup distances are considered and maximizing the smallest distance is used as the criterion for wavelength selection. For quality control purposes, however, it is not important to distinguish among different classes of defective product (or, at most, only secondarily so). The important discrimination is to tell the difference between any defective product and the good product.

It is therefore necessary to modify the basic discriminant criterion so as to optimize the calibration technique to perform the quality control function.

We present the wavelength selection criterion to use for this situation in the form of some actual data where some defective materials were to be distinguished from the good material. In this case the defects in the materials were due to the leaving out of various ingredients that were supposed to be part of the product, as well as not having the correct amount of the desired material present.

Table 15.6 shows the results of an initial wavelength search used to classify good (called *control*) material from several different types of defective material.

TABLE 15.6
Results from Wavelength Search Using All Groups^a

Group	Low	High	Missing A	Missing B	Missing C	Missing D	Missing E
Control	6.5	9.4	3.4	16.8	8.5	9.1	9.9
Low		15.5	9.4	22.3	14.8	14.1	15.7
High			6.5	7.9	1.7	3.9	9.6
Missing A				14.1	6.1	6.5	8.8
Missing B					8.9	8.5	16.5
Missing C						4.8	8.7
Missing D							12.7

^a Wavelengths used for discrimination: 1604, 2248, and 2010 nm.

All the groups were included in the calibration dataset; and the wavelengths selected for this discrimination task were 1604, 2010, and 2248 nm. From the first row in Table 15.6 we see that all the defective materials are satisfactorily distinguished from the control group except group #4: missing A, at a Mahalanobis distance of 3.4 from the control group. Thus it is possible for product without A to be misclassified as being satisfactory, and vice versa. Thus this combination of wavelengths is unsatisfactory for the task at hand.

Examining Table 15.6, we find that the closest pair of groups, the ones with the most nearly similar spectra, are groups #3 (high) and #6 (missing C). While it is this pair of materials that caused the wavelengths to be selected as they were, the wavelengths we wish to find are those that will distinguish all materials from the control group.

Examining the first row of Table 15.6, we see that the missing A group is by far the most difficult one to distinguish from the control group. Thus, to attack the problem, what is needed is a method to distinguish these two materials. In order to find such wavelengths, the sample set was edited so that only the control and the missing A samples were represented in the calibration dataset. Performing a wavelength search on these data only, the wavelengths set — 1870, 2024, and 2080 nm — were selected as the ones to best distinguish the missing A samples from the control samples.

This last triplet of wavelengths was used to perform a discriminant calibration on the full dataset; the results of this calibration are presented in Table 15.7. The discrimination between the control and the missing A groups is improved, as expected, although still not sufficiently to perform a reliable analysis.

Another interesting factor that can be seen in Table 15.7 is that the Mahalanobis distance between the high and the missing C samples has also increased, from 1.7 to 1.9 times Mahalanobis distance. This surprising development might perhaps cast doubt on the reasoning behind the development of the original wavelength selection criterion, but notice that the distance between the high and the missing D samples has dropped to 1.4, a smaller value than in the calibration used for Table 15.6. Thus it was this lack of discrimination that prevented the 1870-, 2024-, 2080-nm set from being selected initially.

The final calibration was generated by retaining the wavelengths found to be best for distinguishing all materials from the control group, the ones used for the results presented in Table 15.7, but separating the control samples from all the others and using only the control samples' data to determine the inverse covariance matrix. This procedure fits the “size” and “shape” parameters optimally to the control group, thus providing the best possible criterion for distinguishing all other materials from it. Table 15.8 shows the results of using this calibration procedure. All

TABLE 15.7
Results from Calibration Using Two Groups to Determine Wavelengths,
but Performing the Calibration Calculations on Whole Dataset^a

Group	Low	High	Missing A	Missing B	Missing C	Missing D	Missing E
Control	5.8	9.1	3.9	15.4	9.0	9.2	7.3
Low		14.8	8.8	21.0	14.8	14.9	12.7
High			6.5	7.0	1.9	1.4	5.5
Missing A				12.7	7.2	6.4	6.7
Missing B					7.8	7.7	12.1
Missing C						2.8	4.4
Missing D							5.3

^a Wavelengths were used: 1870, 2024, and 2080 nm.

TABLE 15.8
Results from Calibration Using Same Wavelengths as in Table 15.7^a

Group	Low	High	Missing A	Missing B	Missing C	Missing D	Missing E
Control	12.5	31.8	22.0	81.0	19.8	27.6	21.0
Low		43.9	33.0	92.6	32.3	39.7	17.8
High			12.5	49.4	14.1	4.5	49.9
Missing A				59.6	12.2	9.1	42.0
Missing B					63.1	53.8	99.2
Missing C						10.5	39.5
Missing D							45.7

^a Inverse covariance matrix calculated from the control group only.

materials representing defective product are now well separated from the control group and are easily distinguishable from it (rather than from each other, an unnecessary criterion).

Applications to Quantitative Analysis. There are two aspects to this application of Mahalanobis distance measures. The first is the direct application of calculating Mahalanobis distances to performing quantitative analysis, and it is discussed in Section 15.2.3.4.1. The second is the calculation of Mahalanobis distances as an adjunct to ordinary regression analysis, and various aspects of this application are discussed in Section 15.2.3.4.2 to Section 15.2.3.4.4.

15.2.3.4.1 Use of Mahalanobis Distances for Quantitative Analysis

This use of multidimensional distance measures is not well developed, but enough work has been done to suggest feasibility [6]. Ordinarily, we consider a set of data to be used for quantitative analysis to be one coherent group, as discussed previously and illustrated in Figure 15.7.

If the large ellipse in Figure 15.7 represents the data from the entire set of specimens, then in this approach we break up this group based on the constituent composition, as shown by the small ellipses in the figure.

Presumably, we already know that there are spectral differences corresponding to the different compositions. Thus, by making each subrange of constituent composition a separate group, classifying a sample as belonging to one or another group is the same as assigning it a value for the constituent concentration: this is, *ipso facto*, quantitative analysis.

As do most things, this new approach has both advantages and disadvantages. One of the potential advantages, for which evidence exists in the reported work [6], is that one group, representing one constituent concentration, can include several different types of material. Thus, this approach may be less sensitive to differences in the ranges of types of material that can be included in one calibration than ordinary regression is.

There are two obvious disadvantages: the first is that it is necessary to have samples in each and every subrange of the constituent range, because each subrange is effectively identified separately. This is a more severe requirement than the corresponding one for regression, which merely requires that the range be covered with a good spread of values.

The second disadvantage is that we lose the auxiliary statistics that the ordinary least-squares regression programs provide, so that it becomes more difficult to ascertain how well the calibration is performing.

15.2.3.4.2 Detection of Outliers in Quantitative Analysis

We have noted in Section 15.2 that data for Mahalanobis distance calculations should have a multivariate normal distribution. A coherent set of data for regression analysis will also meet this

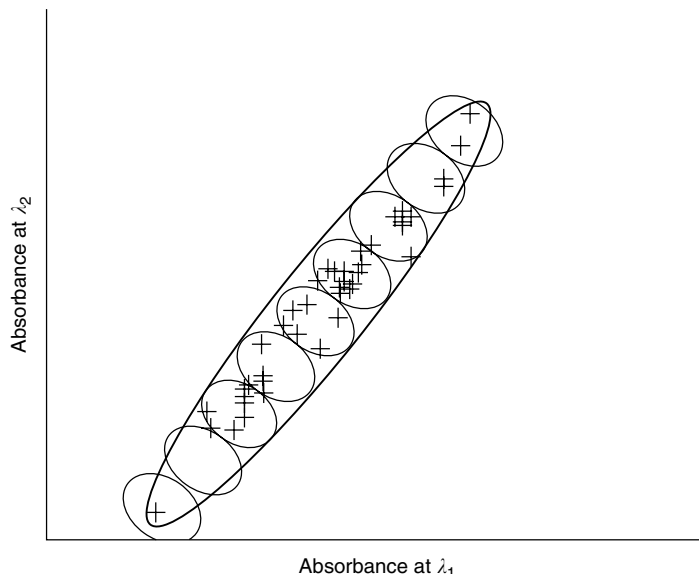


FIGURE 15.7 A method of using Mahalanobis distances to perform quantitative analysis. The large ellipse indicates the region of space occupied by the entire range of the samples; the small ellipses indicate the regions occupied by small subranges of the constituent. By calling each subrange a separate group, an unknown sample can be assigned value for the constituent.

criterion. Thus, fitting a boundary to the region of space occupied by the regression data will allow identification of samples that differ from the training set samples. As shown in Figure 15.5, a sample well outside the boundary of a group will be at a large distance from the group.

One point to note about outliers is that a sample is an outlier if it doesn't fit in with the rest of the samples, whether or not it is predicted correctly. A sample outside the UCL of its corresponding group is such a sample; by measuring the Mahalanobis distance, this case can be identified as such solely on the basis of the optical data alone.

15.2.3.4.3 *Selecting Samples for Quantitative Analysis*

The samples that are optimum for quantitative analysis are those that span the ranges of all variations to which the specimens to be calibrated are subject. In Chapter 8 of this book, the desirability of including samples with all combinations of high and low values of the constituents was noted. However, inclusion of the "normal" range of particle sizes, and any other normal variations of the data; is also desirable, and usually this information is not available.

In the ordinary set of randomly selected specimens that are to be used for calibrating an NIR instrument, most samples are not at the extremes of any of the pertinent variables; if the sample set is large enough, the vast majority of samples will be very close to the group mean, indicating that they are unexceptional in any way. Because most of these specimens occupy the same region of multidimensional space, they are therefore redundant, so that the exercise of collecting the reference laboratory results for them is wasteful of both time and money.

To avoid this waste of time and money, and also to use those samples that have the largest spread of variations not only in those variables whose values are known but also in those variables that we may not even know exist but that have an effect on the spectrum, what is needed is a method of selecting samples that has certain characteristics. Those characteristics are as follows: (a) It is based on measurements of the optical data only, and (b) it selects the samples that show the most differences in the spectra. An algorithm based on Gauss–Jordan reduction has been developed to overcome this problem [7], and here we present an approach based on the Mahalanobis distance concept.

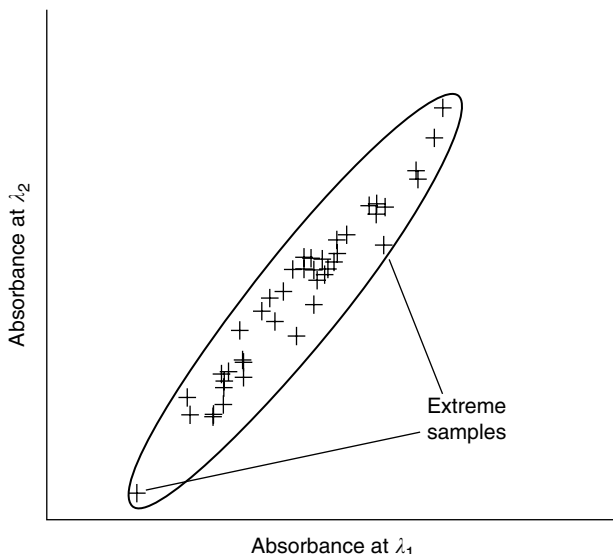


FIGURE 15.8 The data points in different directions are equally far from the group mean in terms of Mahalanobis distances; selecting samples represented by points at the outer fringes of the data cloud results in a set of samples with the maximum variability, and thus optimum for calibration.

The basis of this approach is indicated in Figure 15.8. The data points in the figure represent the optical data that would be collected from a set of calibration specimens. Figure 15.8, as usual, shows the data in two dimensions; but we must keep in mind that in general the concept applies to data in many dimensions. The data are spread out, as is usual, in an ellipsoidal form, representing the multivariate normal distribution of the data. Fitting a boundary ellipse to the data shows that data points are comparably far away from the center of the data, in other words, from the group mean, regardless of their direction from the group mean. However, while the distance from the group mean to a sample's data is determined by the *amount* of variation in the data, the direction from the group mean to a sample's data is determined by the *nature* of the variation. For example, we noted that in Figure 15.2, all the samples appear to lie approximately on a 45° line. This 45° line represents the spread of the data due to repack and particle size effects: Because particle size effects cause data at both wavelengths to increase by nearly the same amount, the effect is to move the data along the 45° line. The spread perpendicular to the 45° line is caused by the variations in the spectra caused by the differing compositions of the specimens measured.

So it is when more than two wavelengths are used: The spread in different dimensions is caused by wavelength changes due to different physical phenomena affecting the spectra. Calculating the shape of the bounding ellipsoid causes all these dimensions to have equal effect in determining how far away from the group mean a given sample lies.

The sample selection algorithm, then, is simply a matter of selecting those specimens that are at the outer fringes of the data cloud. The nature of the process used to determine the shape of the boundary ensures that, if several samples are selected, they will lie in different directions and therefore represent the different sources of variation. Furthermore, the fact that they lie in the fringes of the data cloud ensure that they are the samples that show the maximum amount of variation, as well as the most differences among the causal phenomena.

An advantage of this approach over the Gauss–Jordan method is that, in the Gauss–Jordan method, after a few samples have been selected the residual data represent the noise level and further selections are made essentially at random. In the approach using Mahalanobis distances, samples at the fringes of the data remain there no matter how many samples have already been

selected. Furthermore, the selection process can always be performed in one pass over the data; and there is no uncertainty introduced into the selection process by changing the first sample selected.

15.2.3.4.4 *Evaluating Sample Preparation Methods*

The RMS group size, calculated as described (Section 15.2.2), is also a useful way to select a method of preparing samples for NIR quantitative analysis. Previous work [8] has shown that if a material is subjected to several different methods of preparing it for NIR quantitative analysis, then reading several repacks of each sample prepared by the various methods being tested provides enough information in the optical data alone to determine which sample preparation method is the best. Because the determination will be made solely on the basis of the optical data, it is not necessary to obtain reference laboratory results.

In order to make this determination, obtain enough of one sample of the material for which an ordinary quantitative calibration is desired to provide as many separate repacks as will be needed. Only one sample should be used in order to avoid confusing the effects of the different sample preparation methods with sample variability. For the same reason, the sample must be as homogeneous and well mixed as possible.

Prepare separate aliquots of the sample for NIR analysis using each of the methods that are under consideration. Read each of the different preparations with the NIR instrument, using at least 25 repacks of each preparation. Each set of >25 repacks, each set corresponding to a different preparation method, will be considered a different material and should be labeled as such, in a way that will allow the algorithm to separate the different preparations into different groups.

The interest here is not in determining whether the algorithm can identify which samples were prepared by which method, but rather simply in which method causes the least spread of the data in multidimensional space as measured by the RMS group size (Equation (15.5)). The sample preparation method giving the smallest spread is the one that produces the most homogeneous sample for presentation to the instrument. Thus comparison of the RMS group sizes corresponding to the different sample preparation methods, and selecting the sample preparation method that results in the smallest RMS group size, will provide the best method for quantitative analysis.

All interesting sidelight to the work reported in Reference 8 is the fact that, for a given material, different methods of preparing the sample may be better for quantitative analysis of different constituents.

15.3 METHODS BASED ON USE OF ENTIRE SPECTRUM

For several reasons, performing NIR qualitative analysis based on Mahalanobis distance calculations does not satisfy all needs. One of the main difficulties is the fact that, because matrix operations are involved, it is not practical to perform the necessary calculations on data from more than a relatively small number of wavelengths.

An approach that uses only a small number of wavelengths is also unsatisfactory to many chemists and spectroscopists, who are familiar with comparing spectra visually, and require full spectra, measured at uniformly spaced intervals, for this purpose. In some cases, such as in the pharmaceutical industry, regulatory requirements state that full spectra must be obtained, and also require that spectroscopic methods of identifying unknowns, or even confirming the identity of known materials, must use the full spectrum. In addition, it is necessary to select the wavelengths to use for the identification/discrimination task, and if the starting data are extensive, such as when full spectral scans are available, this can become very time consuming, just as is wavelength selection for quantitative analysis.

The application to NIR spectroscopy has taken two forms to date: direct spectral matching and calculating direction cosines.

15.3.1 USE OF DIRECTION COSINES

This approach to spectral matching was initially developed for use in the UV region of the spectrum [9] and in the mid-IR [10,11] where its use as a matching criterion was applied to the interferograms produced by Fourier Transform (FT) IR spectrometers [12,13] as well as to spectra. Use in the NIR region quickly followed [14,15].

The value calculated as the criterion for deciding how well an unknown sample matches a known is the cosine of the angle between the multidimensional points representing the two spectra. This value has also been termed the “match index” (MI) [14,15] and it calculated from the expression:

$$MI = \frac{\sum_{i=1}^n A_i B_i}{(\sum_{i=1}^n A_i^2 \sum_{i=1}^n B_i^2)^{1/2}} \quad (15.6)$$

The MI takes values between -1 and $+1$; ordinarily a sample is identified as belonging to the group for which the MI is closest to $+1$.

The MI has an interesting geometric interpretation. Figure 15.9 shows the wheat-soy data used previously. In the case of the MI, the quantity calculated is the cosine of the angle between an unknown sample, shown as a cross, and the point representing the known (library) sample belonging to the known group of data, or, as shown in Figure 15.9a, the group mean. If the unknown falls within the bounds representing the given group as shown in Figure 15.9b, it is classified as belonging to that group. Thus, the geometric equivalent of this calculation is essentially a measurement of multidimensional distance using polar coordinates rather than Cartesian coordinates.

Another interesting characteristic of Equation (15.6) is the fact that, if each spectrum in corrected by subtracting the mean value, so each sum of squares becomes the sum of squared differences from the mean $[\Sigma(X - \bar{X})^2]$, then the MI becomes the correlation coefficient between the two spectra.

15.3.2 USE OF DIRECT SPECTRAL MATCHING

This approach calculates the value of the discrimination criterion:

$$\sum_{i=1}^n (X_{1i} - X_{2i})^2 \quad (15.7)$$

where X_1 and X_2 represent the unknown and known (library) spectra.

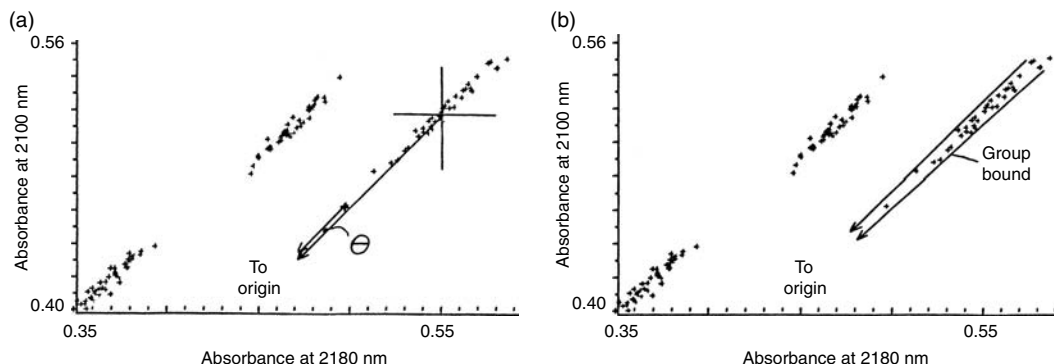


FIGURE 15.9 In polar coordinates the angle between the unknown and the data can be used to classify. If the unknown is within the angular bounds of given group, then it can be assigned to that group.

Use of this approach to the discrimination problem is equivalent to calculating the Euclidean distances between the unknown and each of the known materials; then the sample is identified as belonging to the group to which the distance is smallest.

While this approach has the advantages of both being simple and using the entire spectrum, it is subject to some disadvantages compared to the Mahalanobis distance approach. First, it does not account for the possibility of variations in the compositions of the samples, such as occur in the identification of natural products, or for variations due to physical changes such as particle size/repack effects. Because of the sensitivity to spectral variations caused by compositional changes of the samples, this approach is restricted to the identification of pure chemicals, such as are found in the pharmaceutical and chemical industries. The sensitivity to spectral changes caused by physical variations of the samples means that it cannot be applied directly to the measured spectrum. The spectrum must first be transformed, by mean subtraction, by calculation of the derivative ($dA/d\lambda$), or by some other mathematical transformation that removes the particle size/repack effect from the data.

15.3.3 USE OF PRINCIPAL COMPONENTS

Chapter 8 of this book describes the generation and properties of the mathematical constructs called principal components. Since principal components are calculated solely from the optical data, they are not restricted, as some might think, to quantitative analysis, even though historically their major use in NIR spectroscopy has been for that purpose.

The possibility of using principal components for qualitative analysis has always existed, and recently this use has been described.

The basic concepts of using principal components in this manner are similar to the other methods of computerized qualitative analysis that have already been described in this chapter. After the principal components have been calculated (as described in Chapter 7) the principal component scores are also calculated as described. These scores may now be used as the variables to which any of the described algorithms may be applied.

The principal component scores for each material tend to cluster in multidimensional space, in a similar fashion to the way the data points representing individual wavelength measurements cluster. Thus, any of the methods that are applied to the clusters of data in the “ordinary” qualitative analysis approaches (Mahalanobis distances, direction cosines, etc.) may be used. One approach, which has been used only with the principal component approach to qualitative analysis, has been to demarcate the groups by a maximum (Euclidean) distance that a data point may lie from the group along each principal component axis. This is conceptually equivalent to enclosing the group in a box with planar (or hyperplanar, as appropriate) walls; these walls enclose the boundaries of the group.

The advantages of the principal component approach to qualitative analysis are similar to the advantages of this approach in quantitative analysis: It avoids the need for wavelength selection, and reduces the sensitivity of the classification scheme (whichever one is used) to the random noise of the data.

There is one decision to be made, however, in using this method of qualitative analysis that is not necessary when quantitative analysis is performed: selection of the data to be used in the initial computation of the principal components. There are two possibilities here. One is that the entire dataset obtained from measurements on all the different materials (in the library) to be distinguished may be used together to compute the principal components. Alternatively, the data from each material in the library may be used separately, and separate (and different) principal components calculated for each different material. This latter procedure has been used in other spectroscopic regions, and it is known by the acronym SIMCA. At this time there is no evidence that either approach provides superior performance; indeed, in most practical cases we would expect that the spectral differences are great enough for any of the approaches mentioned to provide essentially perfect identification given any reasonable set of materials in the library. Marginal cases (where the differences in approach

might make a difference in the results) have not yet arisen to the point where they have brought themselves to the attention of the scientific community.

15.4 COMPARISON OF METHODS

Despite the intuitive desire to have an entire spectrum available for identification purposes, it is advantageous to have techniques available that do not require the entire spectrum in order to classify (identify) a substance. Instruments using a finite number of fixed filters, common in the NIR spectral region, cannot use an algorithm based on a complete spectrum. Even when monochromator-based instruments are used, time can be saved by measuring only a small number of preselected wavelengths.

In addition to the limitations of the direct spectral matching approach discussed in Section 15.3.2, there is another, more subtle difficulty that arises when full spectral scans are used for automatic computerized identification of unknown materials. This limitation applies to both of the spectral methods used. To examine the cause of this, let us imagine that we have collected two spectra of the same material, and let us imagine further that we have properly corrected for the effect of particle size shifts on the data. In this case, the two spectra should be essentially identical. However, they will not be actually identical because of the effect of instrumental noise. In this case the effect of noise on the data is easily calculated, and the two spectra will be separated by a multidimensional Euclidean distance equal to σn , where σ is the noise level of the instrument.

In the case of a full spectral scan, n is large; therefore the distance between the spectra will also be relatively large, due solely to the noise of the instrument.

Now let us consider this effect on the discrimination of two spectra that are similar but differ by a small amount at only a few wavelengths. In this case, we would also expect the distance between the spectra to be small, perhaps comparable to the distance between the two identical spectra. Thus, in such a case, the instrumental noise from the wavelengths that do not contribute to the discrimination of the materials blurs the distinction between similar spectra, making them indistinguishable from identical spectra. Thus, when trying to distinguish materials with very similar spectra, we are better off using only those wavelengths that have actual discriminatory power, and leaving out those that do not contribute to the identification process.

Another advantage of the Mahalanobis distance approach is the fact that at this time it is much better understood theoretically. This gives rise to two characteristics: first, it allows theoretical bounds to be specified as the limits for classifying samples and for detecting samples not part of the training set. For both spectral match methods in current use, limiting bounds must be determined empirically.

The second characteristic derived from a well-developed theoretical basis is the wide range of applications, other than strict identification, that have been presented in Section 15.2.3.

A limitation of the use of Mahalanobis distances is encountered when the samples available do not match the conditions for which the Mahalanobis distance concept was generated. If the samples do not exhibit inherent variability within each group, so that a bounding ellipsoid can be generated, then the data cloud collapses into a point in multidimensional space, the size of the cloud being reported as almost zero. Future measurements then report all samples as not having been included in the training set because instrument noise is sufficient to remove the new sample to a sufficiently large distance that, when compared to the calculated size of the groups, is too large to be acceptable. Such a case was recently reported [16]. In that case, the problem was overcome by artificially introducing random noise onto the data in order to enlarge the group sizes. Another solution would be to use Euclidean distances; this would also require an empirical determination of the group boundaries.

Some of the advantages of using full spectra have been discussed in Section 15.3. Another is the savings in calibration time due to not having to perform any wavelength searching. However, this time savings is mitigated during analysis of unknowns because of the longer measurement time needed.

ACKNOWLEDGMENTS

The author gratefully acknowledges permission of Bran and Luebbe Analyzing Technologies to include copyrighted material originally appearing in the discriminant analysis section of the IDAS-PC manual. The author gratefully acknowledges permission of *Applied Spectroscopy* to include the copyrighted material in Table 15.2, originally appearing in *Appl. Spectrosc.* 41: 1204–1213 (1987).

REFERENCES

1. H. L. Mark and D. Tunnell, *Anal. Chem.*, 57: 1449–1456 (1985).
2. H. Mark, *Anal. Chem.*, 58: 379–384 (1986).
3. P. C. Mahalanobis, *Proc. Natl Inst. Sci.*, 2: 49–55 (1936).
4. R. Gnandesikan, *Methods for Statistical Data Analysis of Multivariate Observations*, John Wiley & Sons, New York, 1977, Chapter 4.
5. R. G. Whitfield, M. E. Gerger, and R. L. Sharp, *Appl. Spectrosc.*, 41: 1204–1213 (1987).
6. J. Workman and H. Mark, Paper #121, 14th Annual Meeting of the Federation of the Analytical Chemistry and Spectroscopy Societies, Detroit, MI, October 1987.
7. D. E. Honigs, G. M. Hieftje, H. Mark, and T. B. Hirschfeld, *Anal. Chem.*, 57: 2299–2303 (1985).
8. H. Mark, *Anal. Chem.*, 59: 790–795 (1987).
9. J. A. Reid and E. C. Wong, *Appl. Spectrosc.*, 20: 320–325 (1966).
10. K. Tanabe and S. Saeki, *Anal. Chem.*, 47: 118–122 (1975).
11. P. M. Owens and T. L. Isenhour, *Anal. Chem.*, 55: 1548–1553 (1983).
12. J. de Haseth and L. V. Azarraga, *Anal. Chem.*, 53: 2292–2296 (1981).
13. L. V. Azarraga, R. R. Williams, and J. A. de Haseth, *Appl. Spectrosc.*, 35: 466–469 (1981).
14. K. M. Ronan, P. J. Cooper, and T. H. Begley, Paper #120, 13th FACSS Meeting, St. Louis, MO, September 28 to October 3, 1986.
15. P. J. Cooper, Paper ANYL #111, 193rd National ACS Meeting, Denver CO, April 5–10, 1987.
16. E. Ciurczak, The Discriminant Analysis of Pure Liquids, presented at the 10th International NIRA Symposium, Technicon Science Center, Tarrytown, New York, September 9–10, 1986; abstract available from Bran and Luebbe Analyzing Technologies/Technicon Industrial Systems, Elmsford, New York.

16 Spectral Reconstruction

William R. Hruschka

CONTENTS

16.1 Introduction	333
16.2 Correlation Plots	335
16.3 Difference Spectra	339
16.4 Sharpened Spectra	341
References	344

16.1 INTRODUCTION

Although spectral reconstruction in the near-infrared (NIR) has usually meant the statistical estimation of the shape of the *in vivo* spectrum of a component of a mixture, we use the term in a broader sense to collect here a set of tools that enhance the qualitative understanding of spectra and calibrations. These methods require that “full spectra” be available, in other words, enough closely spaced wavelengths are measured so that a spectrum looks like a continuous curve. The methods should work in any spectral region, but usually require that the measured spectrum be linearly related to the concentrations of the sample components.

We look first at the correlation plot and the related Honigs reconstruction. These use regression at each wavelength to give information about components or correlations. Section 16.3 then discusses difference spectra, which use the difference at each wavelength to give the same kind of information. The last section shows how mathematically sharpening the absorption bands in a spectrum helps perception of the underlying absorbers.

The reader should be warned that these methods do not give definitive answers, but do help one’s understanding, and often guide experimental design. Also, we will see that some refinement of each method is needed to bring out its full utility and avoid mistaken interpretations.

The principles will be illustrated by a set of samples described in Reference 1 and Table 16.1. These are a set of 100 hard red spring wheat samples with a wide range of protein and moisture. The measured protein and moisture allows us to estimate the “starch” concentration as 100–%protein–%moisture, with “starch” remaining in quotation marks because of the method of estimation. The hardness of these samples has been measured by NIR, giving an example in which the component of interest is not well defined and/or has no absorption bands.

For clarity in the plotted spectra, and because this region is available to most instruments described in this book, the spectra from 1600 to 2400 nm are used, although the NIR includes the region 700 to 3000 nm. The spectra were measured at 1.6-nm intervals, giving 500 closely spaced points, and recorded as $\log_{10}(1/\text{reflectance})$ (abbreviated as $\log(1/R)$) satisfying the linearity requirement sufficiently for these examples. Figure 16.1 shows a typical wheat spectrum and the spectra of water, protein, and starch. Figure 16.2 shows four wheat spectra from the illustrative set covering four combinations of high and low protein and moisture. Most of the spectral differences in this figure are caused by particle size or scatter variation from sample to sample. If each spectrum is “normalized” (by an offset and a multiplier) to coincide with the average of the 100 samples at 1680 and 2260 nm,

TABLE 16.1
Composition of 100 Hard Red
Spring Wheat Samples

	Mean	Standard deviation
Water	11.57	1.33
Protein	13.94	2.23
Starch	74.48	2.43
Hardness	96.95	13.82

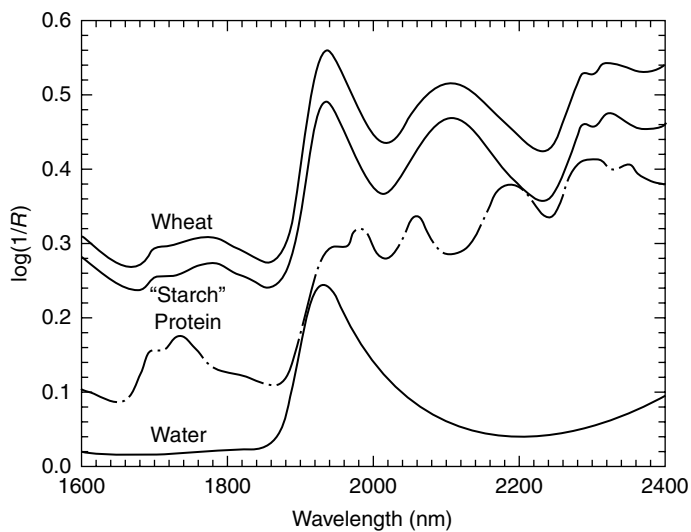


FIGURE 16.1 Spectra of wheat and components.

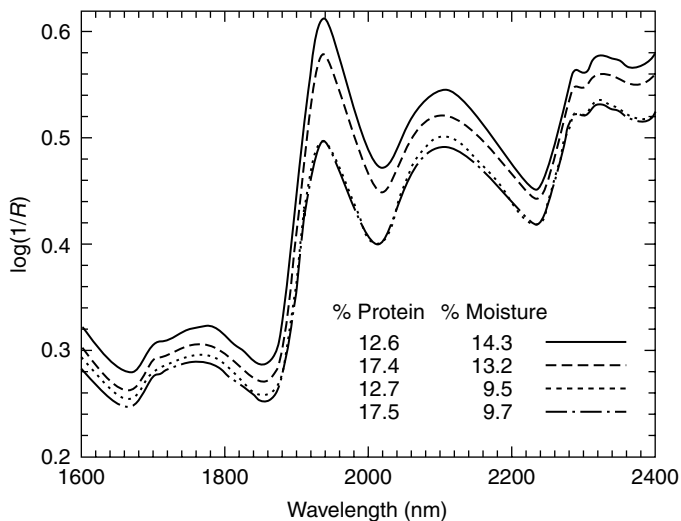


FIGURE 16.2 Four wheat spectra.

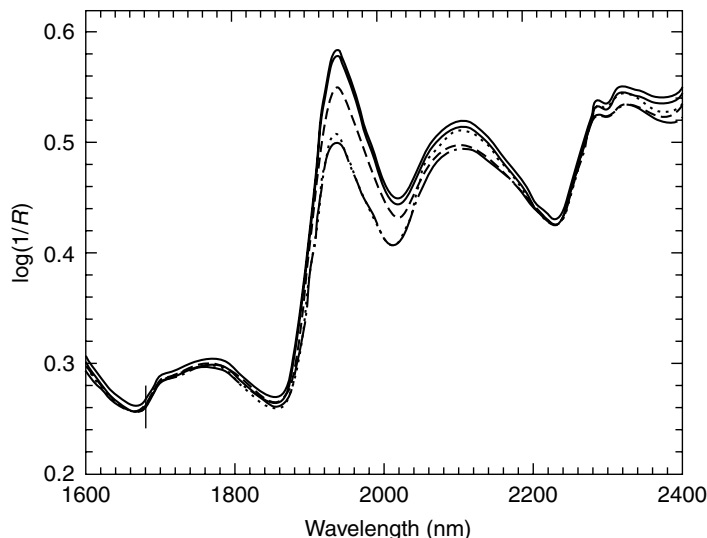


FIGURE 16.3 Spectra of Figure 16.2 normalized to two wavelengths.

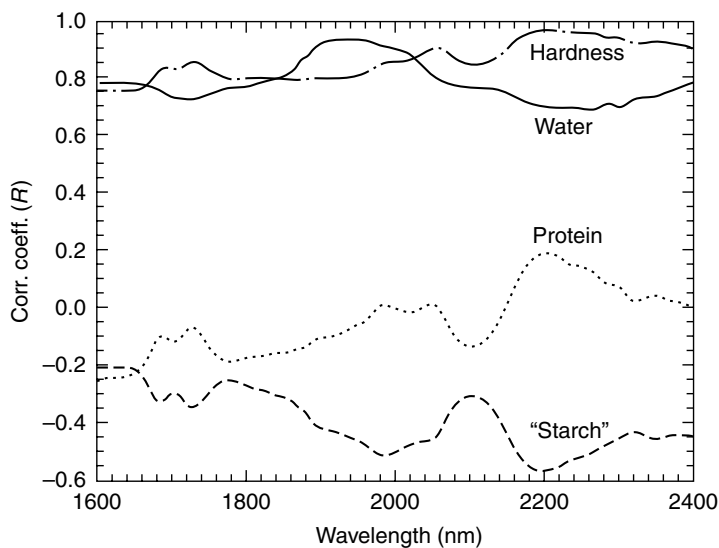


FIGURE 16.4 Correlation plots for wheat components and hardness.

the result in Figure 16.3 allows the compositional differences to predominate. This adjustment could be termed spectral reconstruction because it answers the question, “What would the spectra look like if the scatter effect were the same for all samples?”

16.2 CORRELATION PLOTS

The correlation plot (r plot) is the result of single-term regressions, one at each available wavelength, of composition against $\log(1/R)$. The correlation coefficient (r) is then plotted against wavelength. For example, using the 100 samples, moisture content is regressed against the $\log(1/R)$ value at 1600 nm, giving $r = .7778$. Then the moisture is regressed against $\log(1/R)$ at 1601.6 nm, giving

TABLE 16.2
Correlation between Components of 100 Hard Red Spring
Wheat Samples

	Water	Protein	Starch	Hardness
Water		.02 (<i>r</i> ²)	.17	.35
Protein	-.14 (<i>r</i>)		.71	.16
Starch	-.42	-.84		.48
Hardness	.59	.40	-.69	

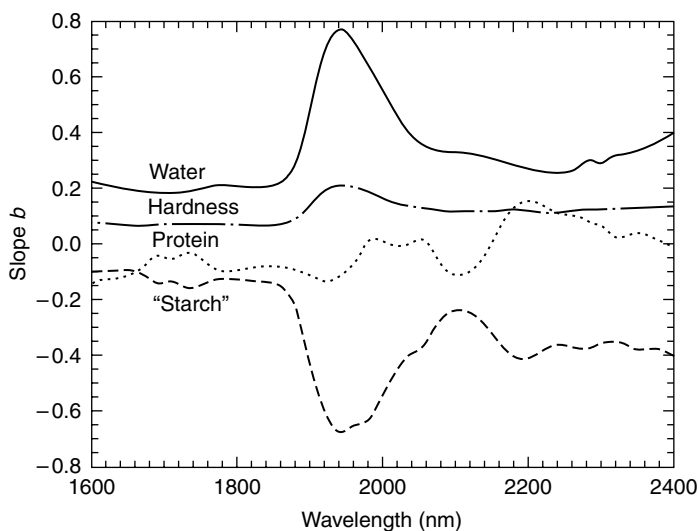


FIGURE 16.5 Slope plots for wheat components and hardness.

$r = .7777$. The regression is repeated at each of the 500 available wavelengths and the r values are plotted as in Figure 16.4, which gives the r plots for water, protein, “starch,” and hardness.

Although the r plots are more useful in derivative or multiterm versions, some information can be had from them: The water and hardness correlations are high at all wavelengths, but the plots have different character. The water plot peaks at 1935, which is the wavelength of the dominant water band. The hardness plot has peaks at the protein peaks, indicating a correlation between hardness and protein. The protein and starch plots are almost mirror images of each other, indicating an inverse correlation between these components. These observations are borne out by the correlations in Table 16.2.

The Honigs reconstruction (H plot) [2] is related to the r plot as follows: each regression (one per wavelength) gives an r and the coefficients in the regression equation:

$$\% \text{ Composition} = a + b \times (\log 1/R)$$

So for each wavelength, we have not only an r , but also a (intercept) and b (slope). Each wavelength also has an average spectral value for the 100 samples. This suggests that the a plot, the b plot, and the average spectrum might also be of interest. These are shown in Figure 16.5 and Figure 16.6. The Honigs method estimates the *in vivo* spectrum of a constituent by adding a multiple of the b plot to the average spectrum. The results for the four example components are shown in Figure 16.7. The water and protein reconstructions show a remarkable similarity to the spectra of the extracted

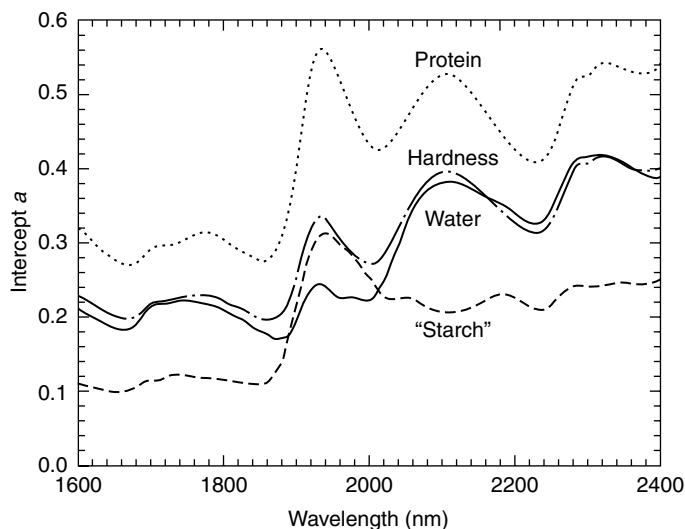


FIGURE 16.6 Intercept plots for wheat components and hardness.

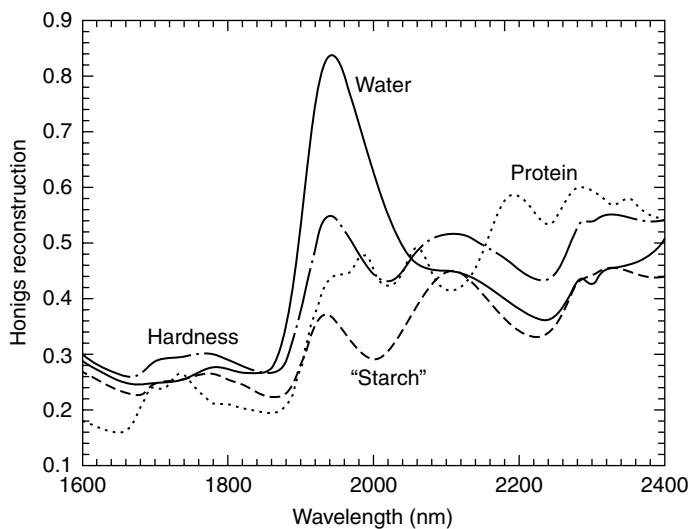


FIGURE 16.7 Honigs reconstruction of the *in vivo* spectra of wheat components and hardness.

components shown in Figure 16.1, as does the “starch” reconstruction, although the starch concentrations were only estimated.

One must be careful when making statements based on these plots. For example, the appearance of water in the protein and starch spectra of Figure 16.1 shows that these pure components had some water in them when their spectra were measured, whereas the appearance of water in the protein and starch reconstruction indicates that there is a correlation between protein and moisture and between starch and moisture in this dataset.

Similarly, in the water H plot, the broad shoulder to the right of the 1940-nm band probably does not indicate some form of bound water, but only a correlation between water and starch in this dataset. But this is not the whole story. Recall that the H plot is a weighted sum of the average spectrum and the b plot. If the average spectrum and the b plot (or r plot) both have peaks at the

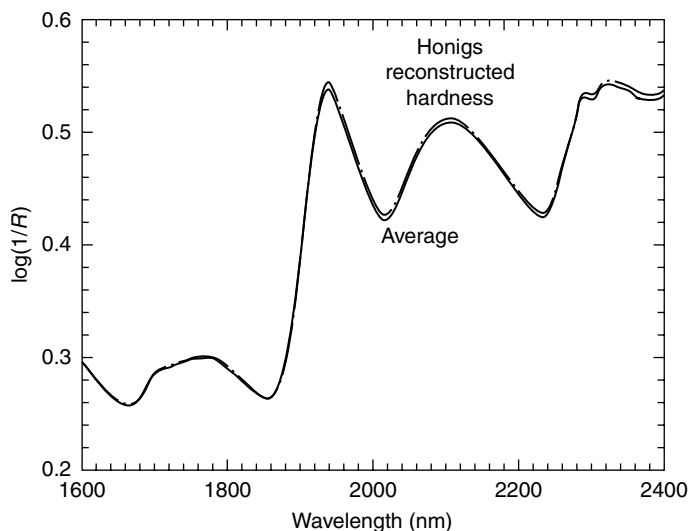


FIGURE 16.8 Average spectra and Honigs reconstructed hardness.

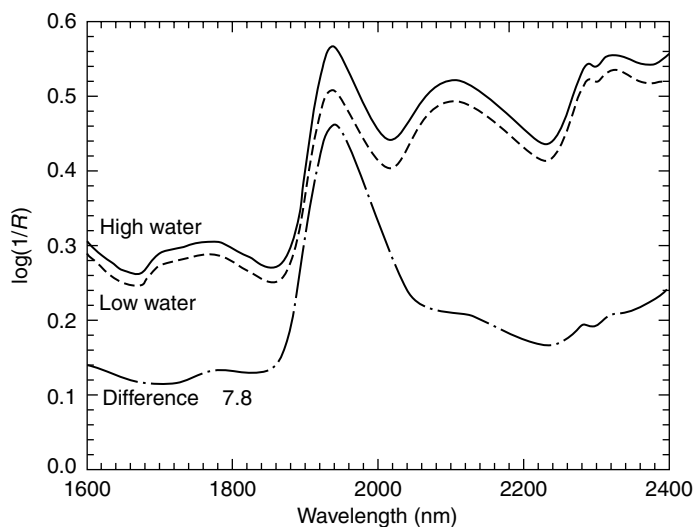


FIGURE 16.9 Difference spectrum between averages of high- and low-water sample sets.

starch band, the H plot must also. It is possible to select a subset of the samples in which there is a negative correlation between water and starch, resulting in an H plot showing less of a starch band.

Figure 16.8 shows that the hardness reconstruction is almost the same as the average spectrum. This is understandable, since the particle size distribution (which is related to hardness) has a similar effect at all wavelengths; and although the correlation to hardness is high at all wavelengths, Figure 16.9 shows that the slope of the regression is low at all wavelengths, resulting in little contribution to the reconstruction. Conversely, if a reconstructed spectrum of a component or characteristic is similar to the average spectrum, one can safely guess that the component is highly correlated to particle size or scattering.

In conclusion, the correlation and related plots are useful, but attention must be paid to the correlations between the components of the samples.

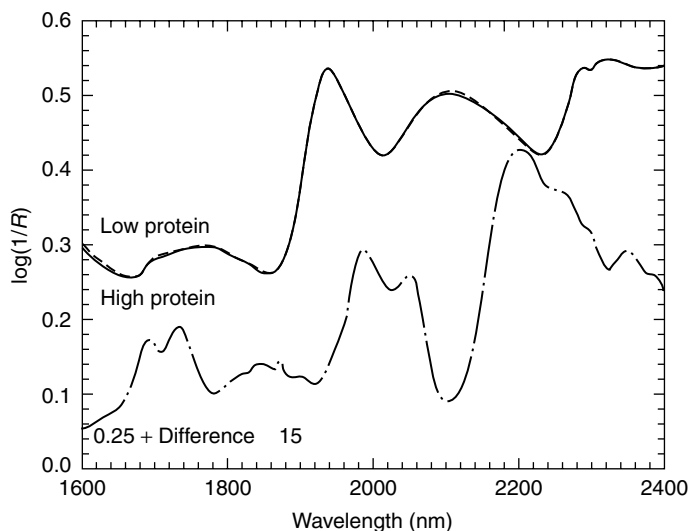


FIGURE 16.10 Difference spectrum between averages of high- and low-protein sample sets.

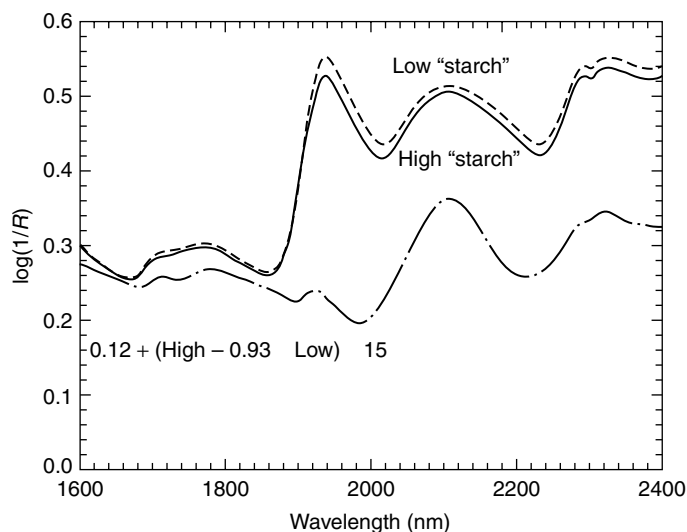


FIGURE 16.11 Difference spectrum between averages of high- and low-starch sample sets.

16.3 DIFFERENCE SPECTRA

Subtracting one spectrum from another, wavelength by wavelength, can give information about component spectra. One must choose the original spectra carefully, and a weighted difference is more useful. Algebraically, if S and T are two spectra, then $D = S - a \times T$ is the weighted difference spectrum. By varying a we can bring out more or less of the desired character. To create Figure 16.9, we divided the 100 samples into two sets — the 50 highest moisture samples and the 50 lowest moisture samples. We then took the average spectrum of each set and took the difference between the average of the high-moisture set and the average of the low-moisture set. Figure 16.10 to Figure 16.12 were made similarly, by ordering the 100 samples by a component, and taking the difference between the average of the high samples and the average of the low samples within that ordering. Table 16.3 gives the corresponding averages of the compositions.

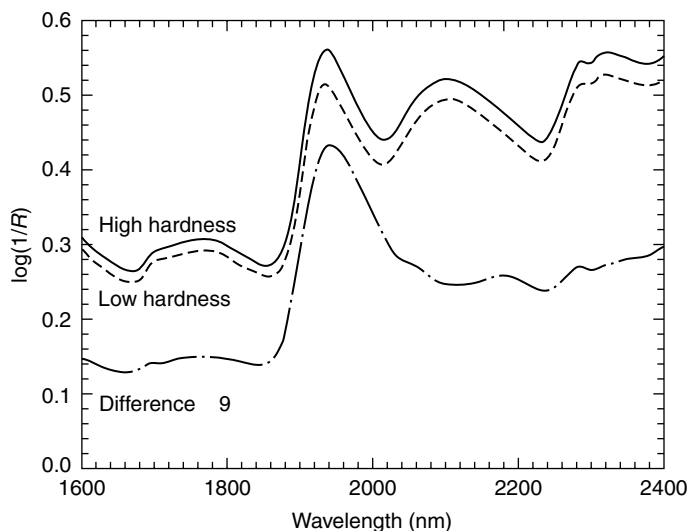


FIGURE 16.12 Difference spectrum between averages of high- and low-hardness sample sets.

TABLE 16.3

Average Composition of 50 High and 50 Low Samples
When 100 Samples Are Ordered by a Component

Ordered by	Water	Protein	Starch	Hardness
Water	<u>12.68</u>	13.61	73.71	104.54
	<u>10.47</u>	14.28	75.25	89.36
Protein	11.47	<u>15.80</u>	72.74	101.41
	11.68	<u>12.09</u>	76.23	92.49
Starch	11.13	12.40	<u>76.47</u>	89.16
	12.02	15.49	<u>72.49</u>	104.74
Hardness	12.22	14.57	73.21	<u>108.18</u>
	10.93	13.32	75.75	<u>85.73</u>

As with the Honigs reconstruction, the similarity of the difference spectra to the expected spectra is remarkable at first glance. But again, correlations between components can lead to misinterpretation. In the water difference spectrum, the shoulder at 2120 nm occurs because not enough of the starch has been subtracted, not because of some previously unknown water absorption band. Table 16.3 shows that the low-water samples average 75.25% starch, which is lower than the 73.71% starch average of the high-water samples. (The direction of this difference is to be expected from the negative correlation between water and starch shown in Table 16.2.) By subtracting more of the low-water spectrum from the high-water spectrum, we can eliminate most of the starch shoulder. Alternatively, as in Figure 16.13, we can choose two samples (not averages of samples) with different water levels, but the same starch/protein ratio. The difference will generally have more noise than the difference between averages, but in this case the false starch peak is small.

Some comments about the protein difference spectrum: First, it has very little water. Second, the small sharp peak at 1870 nm is probably a water vapor band.

Third, the decline in the spectrum after 2200 nm is probably a combination of the lack of water and subtracting too much starch.

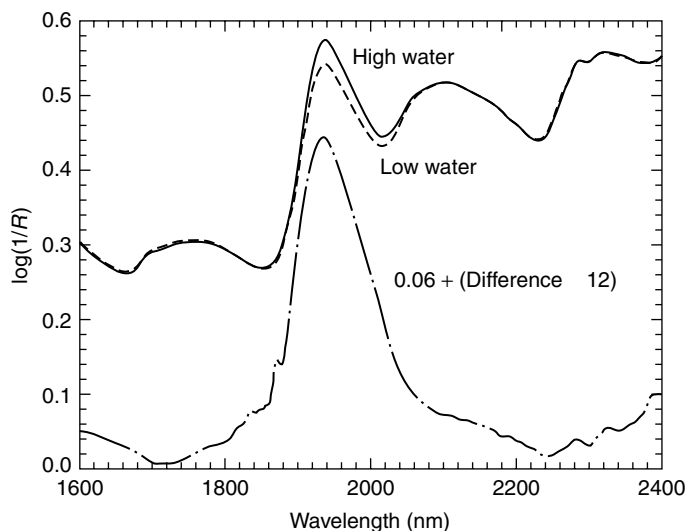


FIGURE 16.13 Difference spectrum between two wheat spectra with the same protein/starch ratio.

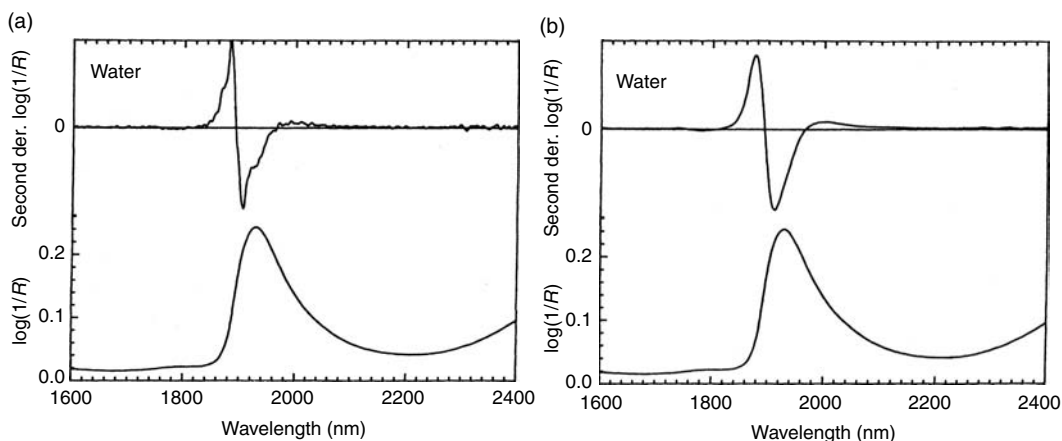


FIGURE 16.14 Spectrum of water and second derivatives. (a) gap = 2.4 nm; (b) gap = 12.0 nm.

16.4 SHARPENED SPECTRA

The NIR spectra of agricultural samples have broad overlapping peaks. There are several methods of mathematically sharpening these peaks so that they do not overlap [3]. Each method requires some trade-off between the amount of sharpening and the creation of artifacts, and usually involves some smoothing of the spectra. We use in these examples the finite second difference (commonly called derivative) method described in Reference 3, preceded by a 4.8-nm running mean smooth. In this method the amount of sharpening is determined by the “gap” parameter. Figure 16.14 shows the sharpening of the water spectrum with a wide (12.0-nm) and a narrow (2.4-nm) gap. Similar spectra can be obtained by other methods by adjusting the corresponding parameters. The derivative has the peculiarity of inverting the spectrum, so that the peaks become narrow valleys.

The corresponding sharpenings of the protein and starch spectra are in Figure 16.15 and Figure 16.16. The fact that the different amounts of sharpening give different numbers of underlying

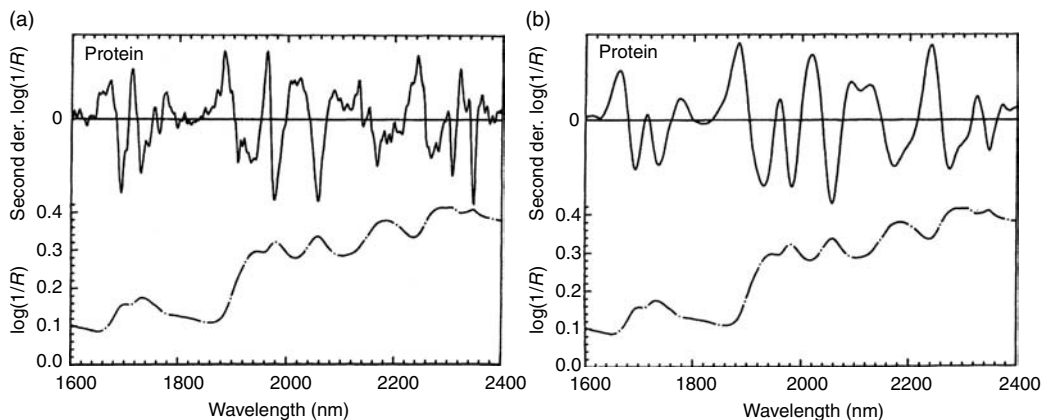


FIGURE 16.15 Spectrum of protein and second derivatives. (a) gap = 2.4 nm; (b) gap = 12.0 nm.

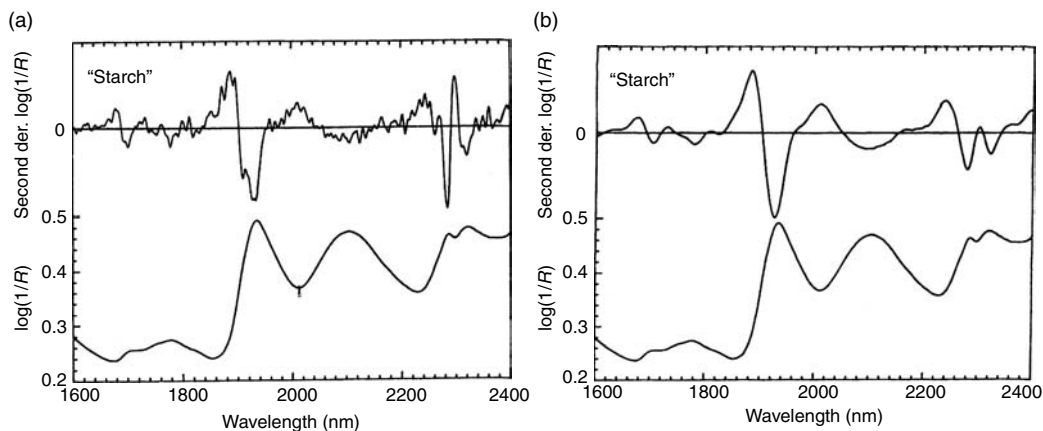


FIGURE 16.16 Spectrum of starch and second derivatives. (a) gap = 2.4 nm; (b) gap = 12.0 nm.

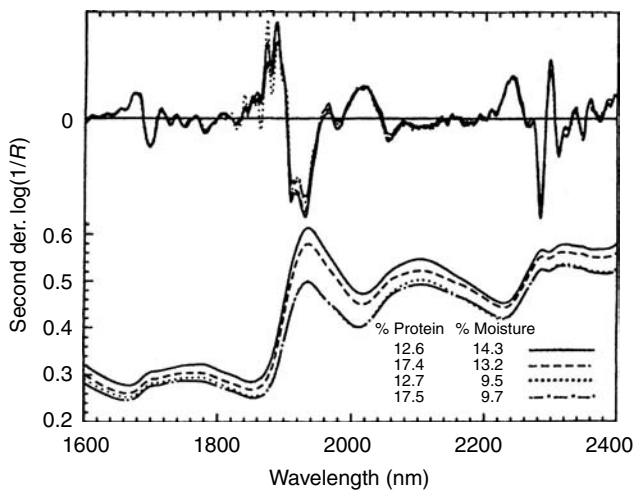


FIGURE 16.17 Spectra of Figure 16.2 and second derivatives (gap = 2.4 nm).

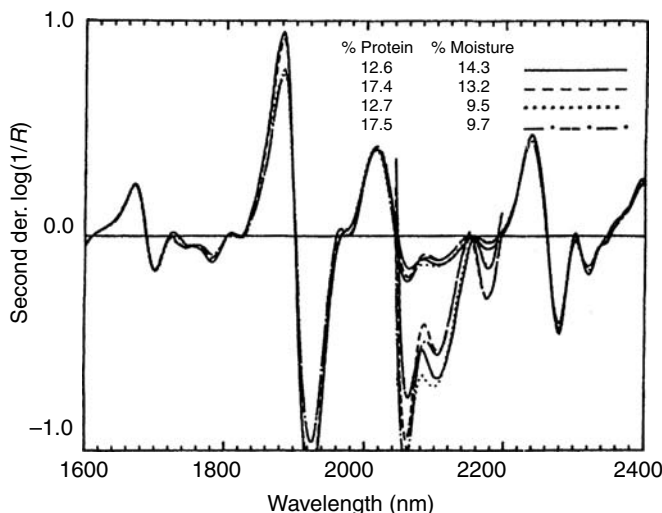


FIGURE 16.18 Second derivatives of spectra of Figure 16.2 (gap = 12.0 nm).

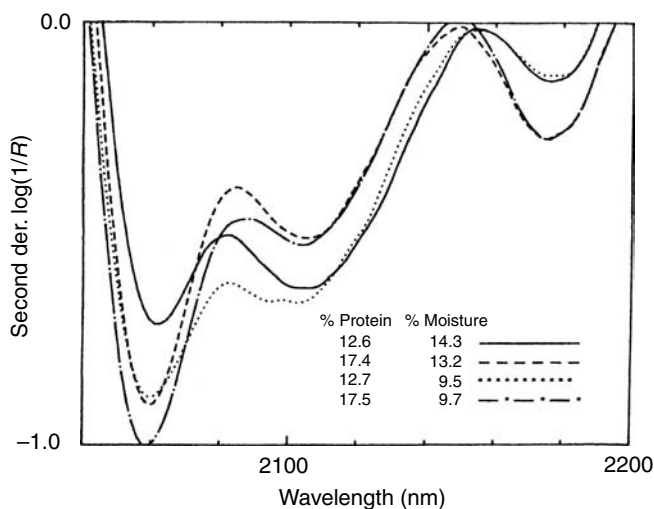


FIGURE 16.19 Detail of Figure 16.17.

peaks tempts one quickly to the question, “How many peaks are *really* there?” This question should be approached carefully. In some situations, as in the spectra of gasses, the absorbers are much narrower than the resolution of the spectrophotometer measuring the spectra. Here we can use sharpening to separate bands that are actually separate in nature. However, with the kind of spectra shown in this chapter, the bands are naturally broad. Sharpening here can introduce an artificial division into subpeaks that has no physical basis, or clever manipulation of the sharpening parameters can produce evidence for a variety of conflicting hypotheses about the underlying absorbers.

Sharpening is a useful tool with naturally broad-banded samples when the goal is to develop calibrations or to understand why calibrations work. We conclude with Figure 16.17 to Figure 16.20 that illustrate how sharpening can accentuate small differences in the spectra of Figure 16.2.

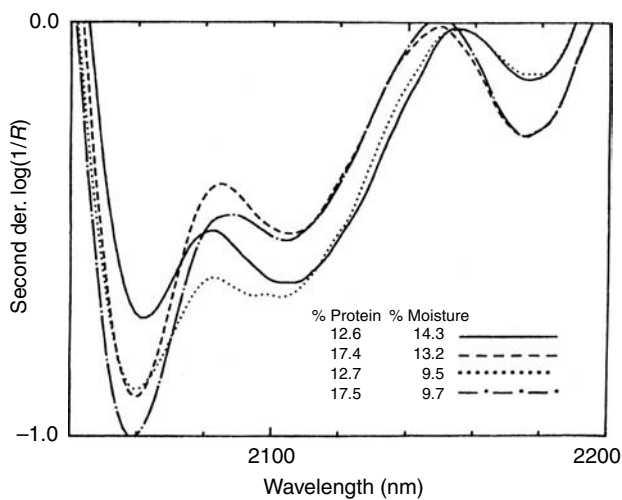


FIGURE 16.20 Detail of Figure 16.18.

REFERENCES

1. W. R. Hruschka and K. H. Norris, *Appl. Spectrosc.*, 36: 261 (1982).
2. D. E. Honigs, G. M. Hieftje, and T. Hirschfeld, *Appl. Spectrosc.*, 38: 317 (1984).
3. W. R. Hruschka, in *Near-Infrared Technology in the Agricultural and Food Industries* (P. Williams and K. Norris, eds.), American Association of Cereal Chemists, St. Paul, MN, 1987, pp. 35–56.

Part IV

Applications

17 Application of NIR Spectroscopy to Agricultural Products

John S. Shenk, Jerome J. Workman, Jr., and Mark O. Westerhaus

CONTENTS

17.1	NIR Technology in Agriculture: 1968 to Present.....	348
17.1.1	Brief History	348
17.2	NIR Spectra of Agricultural Products: Theoretical Aspects	349
17.2.1	Chemical Bonds and NIR Absorption	349
17.2.2	NIR Spectra	351
17.2.3	Derivatized Spectra.....	352
17.2.4	Specific Functional Groups Bands in Agricultural Products.....	353
17.2.5	Special Considerations	355
17.2.5.1	Water and Hydrogen Bonding	355
17.2.5.2	Particle Size Effects in Forage Materials	360
17.3	Reference Methods for NIRS Analysis	360
17.3.1	Definition of the Concept	360
17.3.2	Chemical Reference Methods for Agricultural Products.....	362
17.3.2.1	Moisture	362
17.3.2.2	Protein	362
17.3.2.3	Fiber	363
17.3.2.4	Minerals	365
17.3.2.5	<i>In Vitro</i> and Animal Response	366
17.3.2.6	Conclusion	367
17.3.3	Sample Preparation Methods for NIRS Analyses	367
17.3.3.1	Sampling	367
17.3.3.2	Drying	367
17.3.3.3	Grinding	368
17.3.3.4	Mixing	368
17.3.3.5	Packing	368
17.3.3.6	Unground Samples.....	369
17.4	Calibrating NIR Instrumentation.....	369
17.4.1	Compiling a Sample/Spectral Library	369
17.4.2	Repeatability File	370
17.4.3	Global Calibrations	371
17.4.3.1	Sample Selection	371
17.4.4	LOCAL Calibration	372
17.4.4.1	Samples with Spectral Features	372

17.4.5	Mathematical Transformations of Spectral Data.....	373
17.4.5.1	Quantification	373
17.4.5.2	NIRS Algorithms.....	373
17.4.6	Selecting Calibration Equations	373
17.5	Routine Agricultural Application	374
17.6	Calibration Transfer	375
17.6.1	Basic Problem Defined	375
17.6.2	Methods of Transfer	375
17.6.3	New Methods of Instrument Standardization	377
17.6.4	Calibration Transfer Goal	378
17.6.5	Summary of Major Points	379
17.7	Monitoring NIRS Analyses.....	380
17.7.1	Standard Error of a Difference	380
17.7.2	Control Limits	380
17.7.3	Action	382
17.7.4	Recalibration.....	382
	References	382

17.1 NIR TECHNOLOGY IN AGRICULTURE: 1968 TO PRESENT

17.1.1 BRIEF HISTORY

Agriculture produces the food and fiber needed for human existence. Since the beginning of recorded history agricultural products have been marketed and fed on the basis of their volume or weight (i.e., bushels, tons, etc.). In the past 100 years it became apparent that nutrient content as well as quantity measurements should be considered in deriving proper animal feeding programs. During this 100-year period, laboratory methods were developed and refined to provide nutrient information to the industry; however, nutritional evaluation of these products was and still is expensive and time consuming. Near-infrared spectroscopy (NIRS) analysis offered the promise of rapid low-cost analysis of nutrient composition that could be applied to the ever-expanding requirement for increased efficiency in the feeding of livestock.

The acronym NIRA, or near-infrared analysis, is a term that implies the use of computer algorithms and multivariate data-handling techniques to provide either qualitative or quantitative analysis of a sample (or samples). NIRS includes a single spectral measurement and as such is a more generic definition. For example, an optical engineer involved in the design of a NIR instrument would be involved with NIRS but not necessarily NIRA.

First reports of NIRS were described in the literature as early as 1939 [1,2]. Further work on the potential applications of NIRS to analytical problems was developed by Kaye in 1951 [3–7]. Sutherland suggested in 1950 that with few exceptions, hydrogenic (X–H) stretching vibrations were responsible for all of the absorption bands in the NIR region. Whetsel [8] compiled a review of NIR literature up to 1969. These findings led to the development of commercial NIR moisture monitors in the 1960s.

Karl Norris and coworkers in 1968 [9,10] applied NIRS to the analysis of agricultural products. They recognized the potential of the diffuse reflectance measurement in the NIR region for rapid analysis of grains. These agricultural materials were found to exhibit specific absorption bands in the NIR region. Norris suggested that NIR instruments be used to measure protein and moisture in grains and protein, oil, and moisture in soybeans. The suggested instruments were to contain, at a minimum, the following wavelengths: 1680, 1940, 2100, 2180, 2230, and 2310 nm.

Papers by Norris et al. in 1976 [11,12] added a new dimension to NIRS analysis of agricultural products. This work was conducted on a Cary 14 scanning monochromator and demonstrated that forages could be analyzed by NIRS for quality constituents, including factors such as sheep digestion and intake. Because commercial scanning monochromators were not available, they proposed a new set of wavelengths at 1672, 1700, 1940, 2100, 2180, and 2336 nm for forage materials. This information could be used to construct simple filter instruments for general feedstuff analyses.

By 1977, Shenk et al. [13,14], using a custom-designed spectro-computer system, presented additional evidence that NIRS could provide rapid and accurate analysis of forage quality. In early 1978 they developed portable instrumentation for use with a mobile van that could be taken to the farm and used in hay markets [15,16]. Barton and Burdick [17–19], using a tilting filter instrument, demonstrated that this instrument configuration was capable of analyzing the major quality constituents in forage products acceptably. These findings led to the development of the U.S. Department of Agriculture (USDA) NIRS Forage Network in 1978. One of the major objectives of this network was to develop and test computer software needed for NIRS research in forages. This network of seven laboratories presented a summary of their research findings in USDA Agriculture Handbook 643 in 1985 [20]. The scope of the project was expanded to include grains, soybeans, and other agricultural products, and a new supplement to the handbook was published in 1988.

By 1983 commercial companies began to market NIRS instrumentation and software for forage and feed analysis. With the complex absorptive matrices of these materials, greater wavelength coverage as provided by the scanning monochromators, scanning filter instruments (tilting filters), or fixed filter instruments with more than ten filters became a requirement (Shenk and Barnes, 1977). Simple calibrations with a few filters do not provide acceptable accuracy.

The research prior to 1986 involved instrument and software development, new applications and constituent calibrations, and feasibility studies by university forage extension and instrument manufacturers. The most notable examples of this last point were the extension projects for NIR vans in Pennsylvania, Minnesota, Wisconsin, and Illinois. More recent efforts led to improvements in initial instrumentation, development of improved calibration concepts, standardization of instruments, monitoring systems, and a better understanding of the fundamental aspects of NIRS.

A more recent comprehensive information source for NIR analysis of forages is a recent book edited by C. Roberts, J. Workman, and J. Reeves comprehensively covering nearly all aspects of the use of NIR in agriculture [21]. This work covers traditional forages and feedstuffs as well as multiple applications, a list of which demonstrates the amazing potential of NIR to analyze many commercially important agricultural products. A chapter headings list provides an entrée into the broad application of NIR to agriculture: small grain crops, oil seeds and coffee grains, coffee, tea, spices, medicinal plants, aromatic plants, fruits and vegetables, sugarcane, cereal foods, baking products, beverages, brewing products, fats and oils, dairy, egg, and meat products. In addition, timber and paper, animal by-products, wool, cotton, and soil analysis are included.

17.2 NIR SPECTRA OF AGRICULTURAL PRODUCTS: THEORETICAL ASPECTS

17.2.1 CHEMICAL BONDS AND NIR ABSORPTION

The spectral characteristics of agricultural products are very similar to those of other materials. With the exception of gas phase and high-temperature applications, NIR (0.7–2.7 μm) includes the molecular absorptions of the overtone (0.7–1.8 μm) and combination (1.8–2.7 μm) bands. For each fundamental absorption band there exists a series of overtones with decreasing intensity depending on the increasing multiple or transition number (Table 17.1). The various bands can form a myriad of combination bands with band intensities increasing as frequency decreases (longer wavelengths). NIR band intensities are much weaker than their corresponding mid-infrared (mid-IR) fundamentals by a factor of 10 to 100 (Table 17.1).

TABLE 17.1
Band Intensities for the NIR vs. the Mid-IR Region

Band	Average intensity ($\text{cm}^2/\text{mol.}$)	Relative intensity (%)
Fundamentals	10,000	100
Binary	100	1
Ternary	10	0.1
Quaternary	1	0.01
Quintary	0.05	0.005

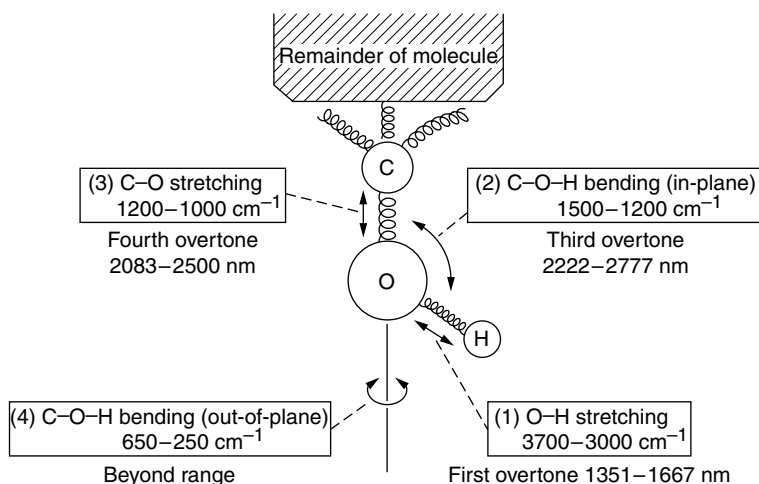


FIGURE 17.1 Vibrations and absorptions of the alcoholic hydroxyl group.

Molecular vibrations exist in the NIR region in the form of X—H, where X is carbon, nitrogen, or oxygen. X—H functionalities are due to hydrogenic stretching, bending, or deformation vibrations. Other important functionalities in the NIR region include the carbonyl carbon-to-oxygen double-bond stretching vibrations, carbon-to-carbon stretching vibrations, and metal halides. The higher wavelengths also contain information on the structural or “skeletal” bending and distortion within a molecule.

Stretching vibrations occur at higher frequencies (shorter wavelengths) than bending vibrations. Stretching vibrations are either symmetrical or asymmetrical whereas bending vibrations are either out of plane or in plane. Out-of-plane bending consists of wagging and twisting, and in-plane bending consists of scissoring and rocking. For any given molecule, stretching occurs at the highest frequency, followed by scissoring, wagging, and twisting and rocking (Figure 17.1). The major bands in the NIR region are second or third harmonics of fundamental O—H, C—H, and N—H stretching vibrations found in the mid-IR region. Monochromatic light produced by an NIR instrument interact with finely ground plant material as reflection, refraction, absorption, diffraction, and transmission (Figure 17.2). Loss of energy from the sample can occur due to specular reflection, internal refraction, complete absorption, and trapping losses due to wide-angle solid ray reflection. If the sample absorbs none of the incident energy, total reflection occurs. Agricultural products selectively absorb NIR radiation that yields information about the molecular bonds within the material being measured. Although Beer’s law, which is summarized in its simplest form by the statement that molecular bond concentration is linear with $\log 1/\text{transmission}$ is assumed valid for all NIR spectroscopy; we note that in fact there is no definitive theory for diffuse reflectance.

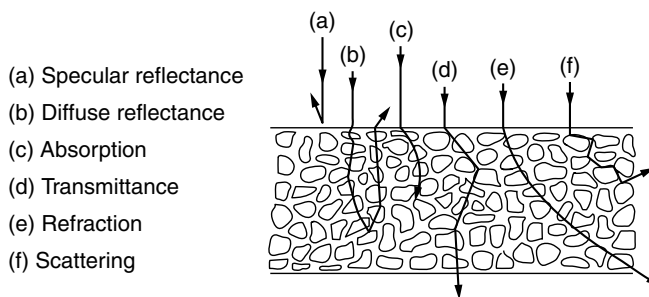


FIGURE 17.2 Interaction of NIR radiation with solid particles in a sample.

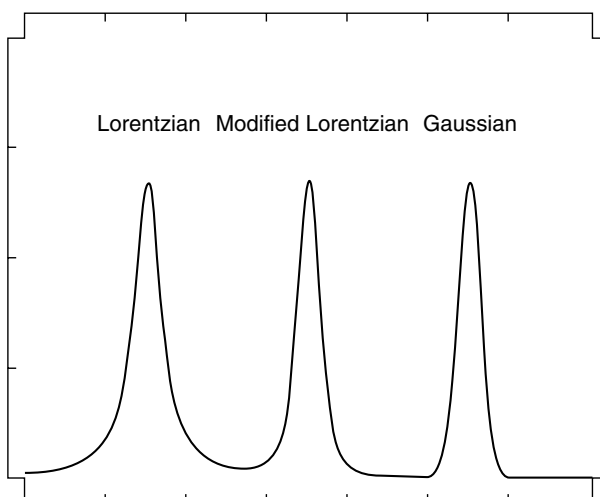


FIGURE 17.3 Simulated absorption bands in Lorentzian, modified Lorentzian, and Gaussian waveforms.

17.2.2 NIR SPECTRA

An NIR absorption band is produced when NIR radiation at a specific frequency vibrates at the same frequency as a molecular bond in the sample. The theoretical shape of the absorption band (Figure 17.3) is Lorentzian in the frequency (wavenumber) scale. The slit shape of the instrument gives a smoothing function to the band that approaches a normal distribution, termed the *slit distribution function*. The final shape of the band is somewhere between Lorentzian and Gaussian (normal). An absorption peak (Figure 17.3) is defined as the highest absorption in a band. Shoulders are noticeable deviations in the spectral shape that do not have a defined maximum. NIR data points are usually collected from a sample in log of inverse reflectance ($\log(1/R)$) form. Resulting spectra contain only a few definable absorption peaks (Figure 17.4).

Bands are defined by three parameters — location, height, and width — and cannot be accurately estimated in $\log(1/R)$ form. Valleys are created between peaks and have different shapes due to adjacent and overlapping bands with different heights and widths. The estimate of band location is probably the most accurate measurement of these three parameters.

The estimates of band height and width are difficult to make because of the overlapping of nearby absorption bands. In overlapping bands, an increase or decrease of intensity in one of the bands can “appear” as a band shift or change in location for the remaining overlapped bands. $\log(1/R)$ is the preferred form of collecting NIR spectra. In this relationship absorption is assumed linear

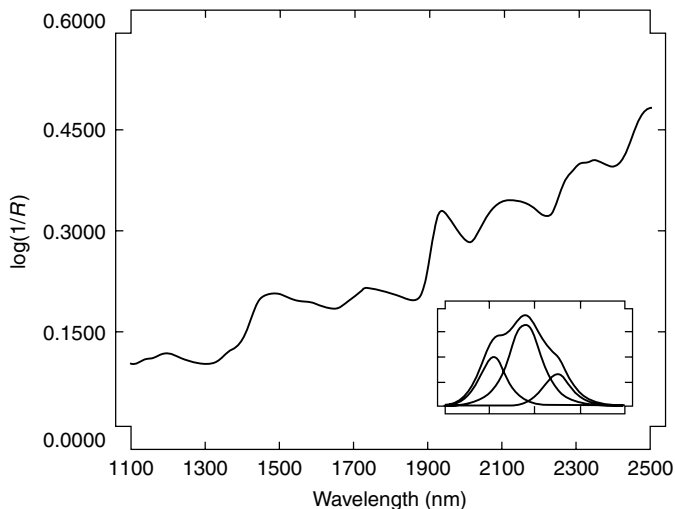


FIGURE 17.4 Forage spectrum displayed in $\log(1/R)$ form with inset showing simulated absorption bands.

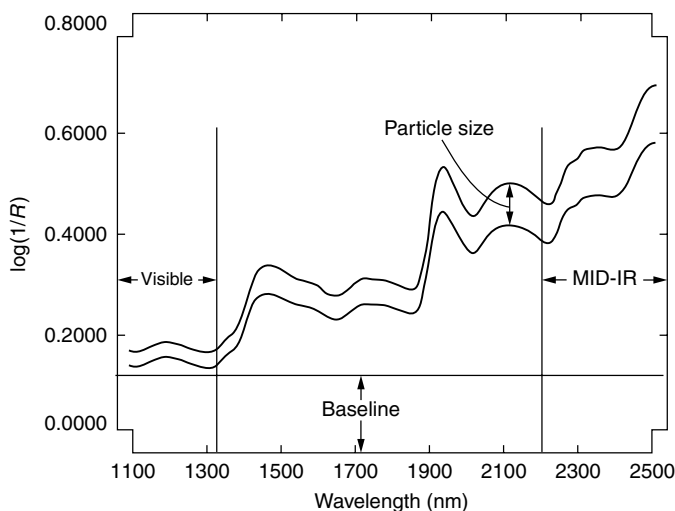


FIGURE 17.5 General features of a NIR spectrum of an agricultural product.

with concentration. A typical spectrum of an agricultural product contains seven to ten peaks with numerous shoulders. These peaks are composites of numerous individual bands that cannot be visibly resolved in $\log(1/R)$ form. Figure 17.4 is a typical forage spectrum with an inset simulating possible underlying peaks in a small portion of the curve.

General features of a spectrum include a multiplicative response to changes in particle size (Figure 17.5), confounding of the 1100- to 1400-nm region with information from the visible region, specular reflectance that is proportional to absorption and therefore greater at the shorter wavelengths, confounding of the 2300- to 2500-nm region with mid-IR tails trailing into NIR, and shifting of the baseline caused by differences in sample holder glass thickness or placement of the sample.

17.2.3 DERIVATIZED SPECTRA

The spectra of agricultural products are often displayed in difference or derivatized form. One form of expressing derivative math treatments is D , G , $S1$, and $S2$ where D is the derivative number

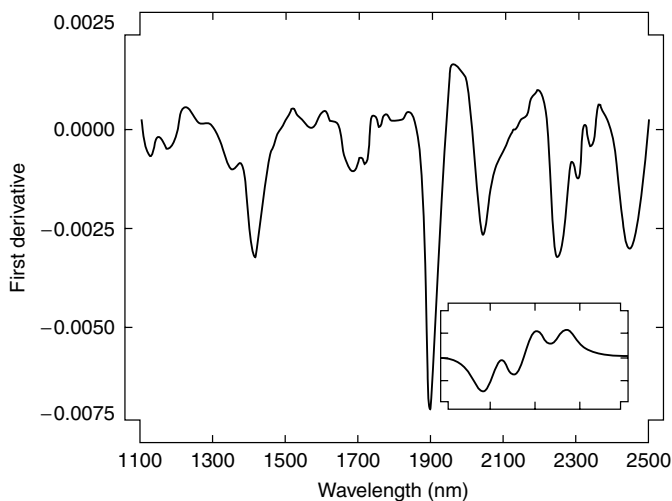


FIGURE 17.6 Forage absorbance spectrum displayed as a first-order derivative with simulated absorption bands.

(order), G is the gap between points used to calculate the difference, and $S1$ and $S2$ are the number of data points used to smooth the data. A first-order derivative of $\log(1/R)$ results in a curve containing peaks and valleys that correspond to the point of inflection on either side of the $\log(1/R)$ peak (Figure 17.6). The simulation figure in the lower right corner of the figure demonstrates this feature. It is rather difficult to visually interpret first derivative because band peaks and valleys do not follow the $\log(1/R)$ spectral pattern.

The second-order derivative calculation results in a spectral pattern display of absorption peaks pointing down rather than up. Second derivative can be very helpful in spectral interpretation due to the fact that in this form band intensity and peak location are maintained with those in the $\log(1/R)$ spectral pattern, and apparent band resolution enhancement takes place. The gap size and amount of smoothing used to make the transformation will affect the number of apparent absorption peaks. Using a small gap size with second derivative math treatment (2, 2, 2, 1) will produce 35 to 40 peaks in the spectrum of most agricultural products. The major advantage here is that second derivative generates few if any false peaks in the negative direction. However, second derivative generates two false valleys in the positive ordinate scale for every band in a negative direction. This effect on the spectral pattern can be seen from the simulated example included with Figure 17.7.

The third-order derivative is rarely used to interpret spectra. It exhibits the same qualities as a first-order derivative and is even more difficult to interpret meaningfully. The fourth-order derivative, on the other hand, is a very useful math treatment to aid in spectral interpretation. A very narrow gap (4, 4, 3, 1) will generate 60 to 70 apparent absorption peaks (Figure 17.8) pointing up. Peak location becomes clearer in this mode because little overlap is experienced. This math treatment emphasizes narrow absorption bands while broad peaks with 80- to 100-nm widths may be difficult to see. The fourth derivative can provide interesting new information but might cause problems with spectral interpretation due to mathematical artifacts such as false valleys and side lobe bands, as is shown by the simulated example in Figure 17.8.

17.2.4 SPECIFIC FUNCTIONAL GROUPS BANDS IN AGRICULTURAL PRODUCTS

As previously stated, absorption bands are produced when NIR radiation at specific frequencies (wavelengths) resonates at the same frequencies as a molecular bond in the sample. Molecular bonds are often described as simple harmonic oscillators. The NIR absorptions are referred to as bands

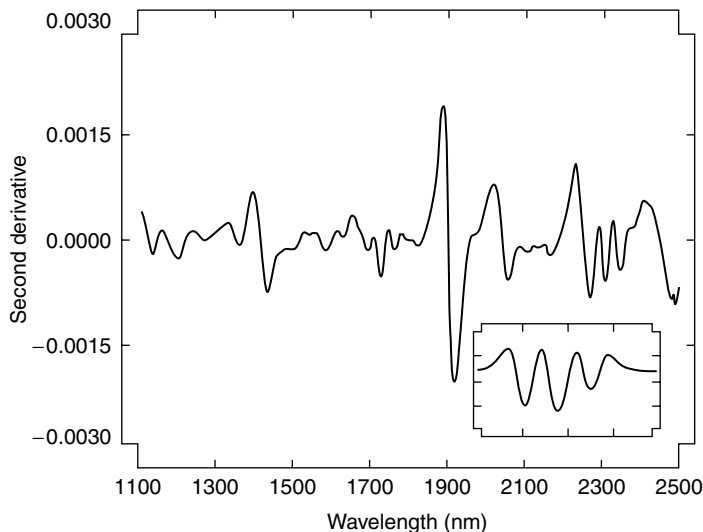


FIGURE 17.7 Forage absorbance spectrum displayed as a second-order derivative with an inset showing simulated absorption bands.

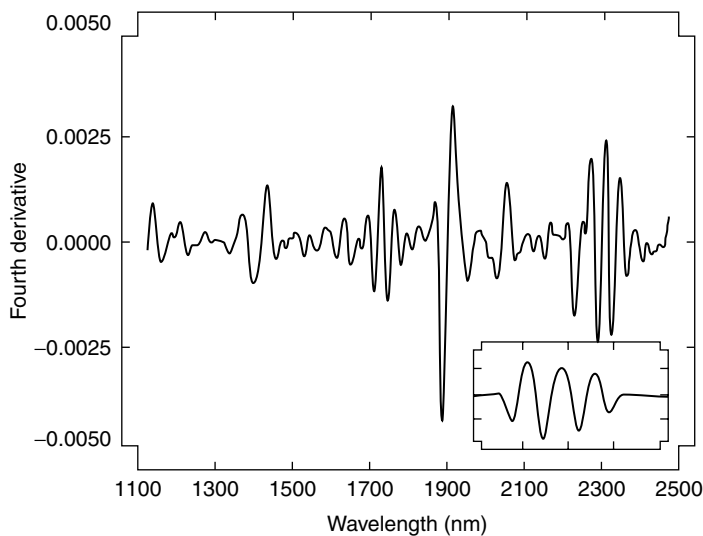


FIGURE 17.8 Forage absorbance spectrum displayed as a fourth-order derivative with an inset showing simulated absorption bands.

because even a monochromator does not produce radiant energy at specific wavelengths but rather as a band that is usually 10 to 12 nm wide in dedicated NIR instruments.

The NIR analyst is most interested in the major molecular bonds found in whole plant material. Figure 17.9 was developed to give a general picture of where the X—H vibrational information can be found. To date, optimum wavelengths for NIRS analysis have relied on empirical calibrations to predict quality constituents for agricultural products. The difficulty in preselecting the best set of wavelengths to use for any particular analysis is brought about by the broad array of chemical compounds present in the sample. This results in extensive overlapping and perturbed NIR absorption bands. The term *wavelength* becomes less useful in this situation because although it relates to

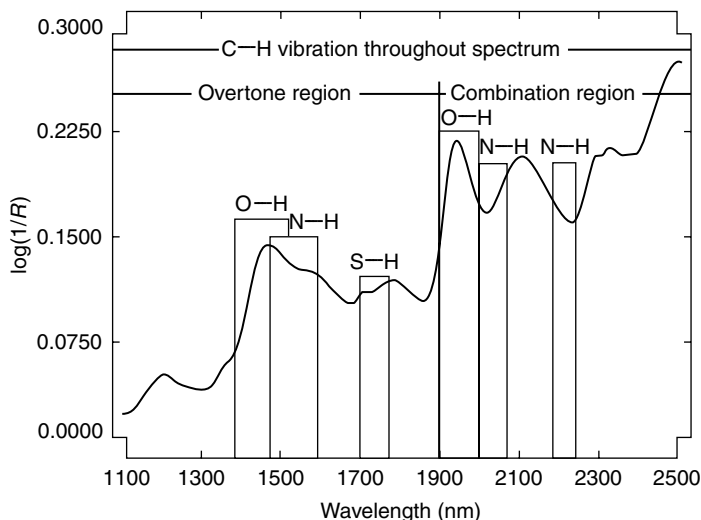


FIGURE 17.9 General representation of X—H vibrational information available in agricultural products.

a specific band location in the spectra, the band itself is a composite of many bands containing information on more than one type of molecular vibration. Another difficulty in relating spectra to specific functionalities is found in the fact that reference methods used in calibration do not measure vibrational information but rather chemical or physical properties that do not have simple molecular vibrational explanations.

Although not complete, data have been compiled on band positions, ranges, intensities, and types of characterizing locations in agricultural products using a variety of sources [22–24]. Basic information exists in the literature that indicates the frequencies (wavelengths) of bond absorptions occurring for generic functional groups as well as those functionalities most important to the NIR analyst [25–27] (Table 17.2).

17.2.5 SPECIAL CONSIDERATIONS

17.2.5.1 Water and Hydrogen Bonding

The transmittance spectrum of pure water is apparently rather simple in $\log(1/R)$ form. When transformed into a fourth derivative, it becomes more complex with a number of apparent major and minor absorption peaks (Figure 17.10). It is possible that some of the apparent peaks are false because of fourth-derivative side lobe artifacts, but it is obvious that neither the first overtone band at 1450 or the combination band at 1930 can be generated from a single Gaussian or Lorentzian absorption band. Each of these broad bands at 1450 and 1930 contains information on more than one hydrogen-bonded subspecies. It is known that variations in hydrogen-bonded molecular subspecies can cause band broadening and peak position shifts. Variations in hydrogen bonding (intra- and intermolecular) result in changes in the force constants of the X—H bonds. The largest change in the force constants of the fundamental vibrations occurs for the X—H stretching vibration. It is known that in the case of O—H vibrations, the apparent band location is isolinear and changes position linearly with changes in temperature.

This information has important implications in agricultural products. The water bands at 1450 and 1930 nm consist of multiple overlapping bands. The apparent location of these composite bands changes as the spectra are measured from one sample type to another. This apparent shift in band position is due to the changes in the relative proportions of the individual bands making up the composite bands. In addition, the concentrations of the molecular subspecies in any particular sample

TABLE 17.2
Basic Characterizing Wavelengths in the NIR Region^a

Wavelength (nm)	Characteristic
2270	Lignin
2310	Oil
2230	Reference
2336	Cellulose
2180	Protein
2100	Carbohydrate
1940	Moisture
1680	Reference

Wavelength	Bond vibration	Structure
1143	C—H second overtone	Aromatic
1160	C=O stretch fourth overtone	C=O
1170	C—H second overtone	.HC=CH
1195	C—H second overtone	.CH ₃
1215	C—H second overtone	.CH ₂
1225	C—H second overtone	.CH
1360	C—H combination	.CH ₃
1395	C—H combination	.CH ₂
1410	O—H first overtone	ROH
		Oil
1415	C—H combination	.CH ₂
1417	C—H combination	Aromatic
1420	O—H first overtone	ArOH
1440	C—H combination	.CH ₂
1446	C—H combination	Aromatic
1450	O—H stretch first overtone	Starch
		H ₂ O
1450	C=O stretch third overtone	C=O
1460	Sym N—H stretch first overtone	Urea
1463	N—H stretch first overtone	.CONH ₂
1471	N—H stretch first overtone	CONHR
1483	N—H stretch first overtone	.CONH ₂
1490	N—H stretch first overtone	CONHR
1490	O—H stretch first overtone	Cellulose
1490	Sym N—H stretch first overtone	Urea
1492	N—H stretch first overtone	ArNH ₂
1500	N—H stretch first overtone	.NH
1510	N—H stretch first overtone	Protein
1520	N—H stretch first overtone	Urea
1530	N—H stretch first overtone	RNH ₂
1540	O—H stretch first overtone	Starch
1570	N—H stretch first overtone	CONH
1620	C—H stretch first overtone	=CH ₂
1685	C—H stretch first overtone	Aromatic
1695	C—H stretch first overtone	.CH ₃
1705	C—H stretch first overtone	.CH ₃
1725	C—H stretch first overtone	.CH ₂
1740	S—H stretch first overtone	—SH
1765	C—H stretch first overtone	CH ₂
1780	C—H stretch first overtone	Cellulose

TABLE 17.2
(Continued)

Wavelength	Bond Vibration	Structure
1780	C—H stretch/HOH deformation combination	Cellulose
1790	O—H combination	H ₂ O
1820	O—H stretch/C—O stretch second overtone combination	Cellulose
1860	C—C1 stretch sixth overtone	C—C1
1900	C=O stretch second overtone	—CO ₂ H
1908	O—H stretch first overtone	P—OH
1920	C=O stretch second overtone	CONH
1930	O—H stretch/HOH deformation combination	Starch
		Cellulose
1940	O—H bend second overtone	H ₂ O
1950	C=O stretch second overtone	—CO ₂ R
1960	O—H stretch/O—H bend combination	Starch
1980	Asym N—H stretch/amide II ^b combination	CONH ₂
1990	N—H stretch/N—H bend combination	Urea
2030	C=O stretch second overtone	Urea
2050	N—H/Amide II ^b or N—H/Amide III ^b or combination	CONH CONH ₂
2055	Sym N—H stretch/amide I ^b combination	Protein
2060	N—H bend second overtone or N—H bend/N—H stretch combination	Protein
2070	N—H deformation overtone	Urea
2070	O—H combination	Oil
2090	O—H combination	.OH
2100	O—H bend/C—O stretch combination	Starch
2100	Asym C—O—O stretch third overtone	Starch or cellulose
2140	C—H stretch/C=O stretch combination or sym C—H deformation	? Oil
		NC=CH
2170	Asym C—H stretch/C—H deformation combination	HC=CH
2180	N—H bend second overtone	Protein
	C—H stretch/C=O stretch combination	Protein
	C=O stretch/amide III ^b combination	Protein
2200	C—H stretch/C=O stretch combination	—CHO
2270	O—H stretch/C—O stretch combination	Cellulose
2280	C—H stretch/CH ₂ deformation	Starch
2300	C—H bend second overtone	Protein
2310	C—H bend second overtone	Oil
2322	C—H stretch/CH ₂ deformation combination	Starch
2330	C—H stretch/CH ₂ deformation combination	Starch
2335	C—H stretch/C—H deformation	Cellulose
2352	CH ₂ bend second overtone	Cellulose
		Protein
2380	C—H stretch/C—C stretch combination	Oil
2470	C—H combination	.CH ₂
2470	Sym C—N—C stretch first overtone	Protein
2488	C—H stretch/C—C stretch combination	Cellulose
2500	C—H stretch/C—C and C—O—C stretch	Starch
2530	Asym C—N—C stretch first overtone	Protein

^a Original work.^b Amide I: C=O stretch. Amide II: N—H in-plane bend; C—N stretch. Amide III: N—H in-plane bend; C—N stretch.

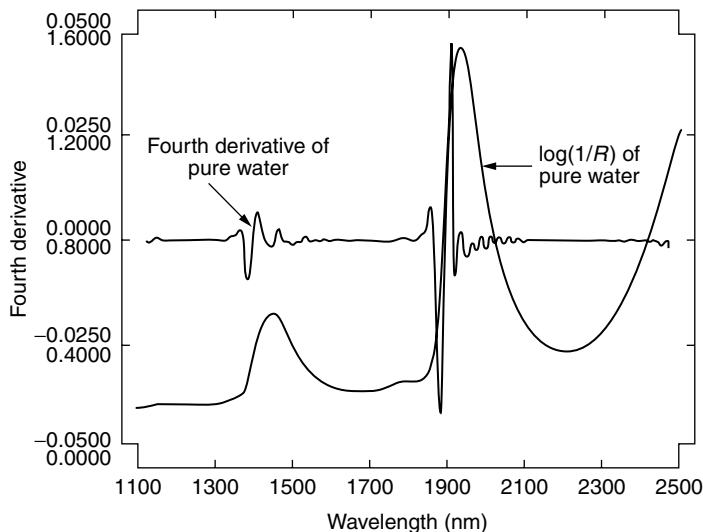


FIGURE 17.10 Spectrum of pure water in transmittance form displayed in $\log(1/R)$ and fourth-order derivative.

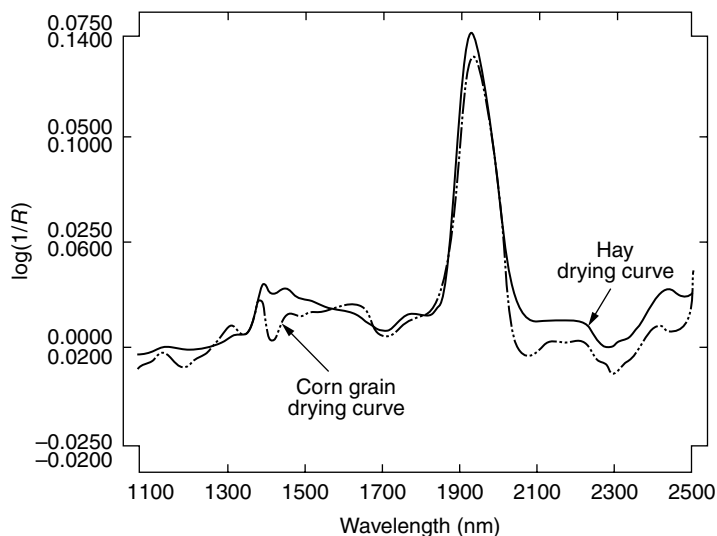


FIGURE 17.11 Difference spectrum of forage and corn grain samples before and after drying.

are directly influenced by chemical interactions with other molecular species in the sample. The largest observable changes in the NIR region are due to changes in hydrogen bonding. To demonstrate this relationship, a difference spectrum was obtained for two agricultural products before and after drying. Although this difference spectrum is not a perfect representation of the O—H bonding in the product, it does provide a profile for comparing changes in O—H band positions and shapes for different materials (Figure 17.11). Not only do these O—H patterns differ from one agricultural product to another, but these profiles can be altered easily by changing sample temperature. As the sample is cooled the individual band at 1900 becomes reduced and the individual band at 1928 increases. This causes a shift in the composite 1930 band and is caused by changes in H bond formation and/or H bond breakage. These small changes can have serious effects on analytical accuracy. This principal is demonstrated in Figure 17.12 and Figure 17.13 with a fourth-derivative

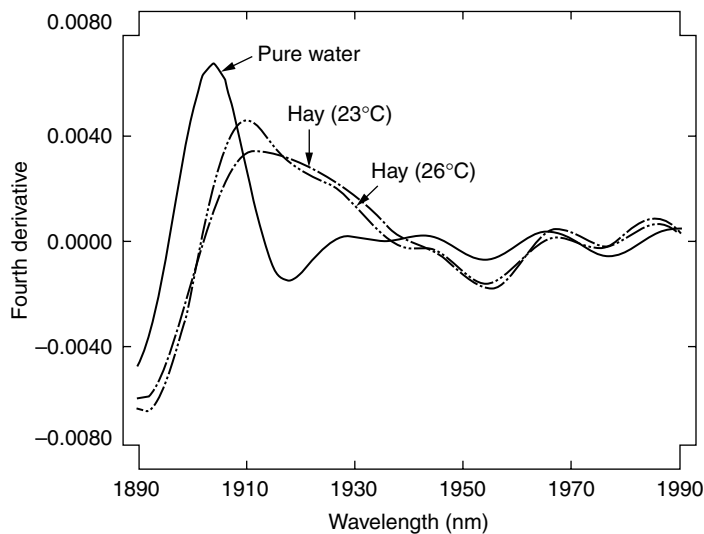


FIGURE 17.12 Influence of a 3°C change in room temperature. Hay absorbance spectra displayed as fourth-order derivatives with a fourth-order derivative transmittance spectrum of pure water included for reference.

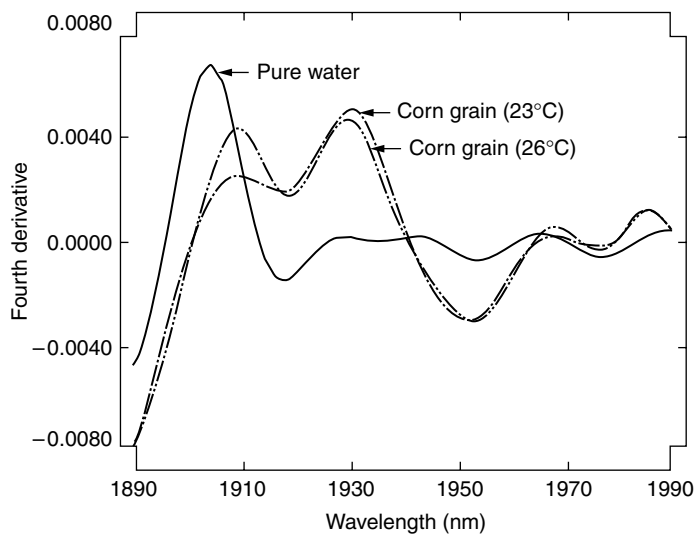


FIGURE 17.13 Influence of a 3°C change in room temperature. Corn grain absorbance spectra displayed as fourth-order derivatives with a fourth-order derivative transmittance spectrum of pure water included for reference.

transformation of a forage and grain sample differing in only 3°C. The fourth derivative of pure water has been added to show the general agreement between O—H band location of these individual bands in pure water and forage and grain.

Hydrogen bonding and chemical equilibrium relating to hydronium ion concentration can be determined. Increased mass upon hydrogen bonding due to inter- and intramolecular interaction can be determined by O—H bands shifting to higher wavelengths and by band broadening or narrowing, and even by apparent band shape and intensity changes. If there is an increase in non-hydrogen-bonded species within a sample, one might observe several effects on the spectral pattern, including band position shifts to lower wavelengths and band narrowing. The hydrogen bonding relating

to sample chemistry must be considered in high-moisture forage samples and especially in slurry samples containing greater than about 50% water. These spectra must be carefully monitored in terms of temperature, ionic strength, and other factors relating to hydrogen-bonded species in solution.

To understand shifts in the frequency of hydrogen-bonding vibrations we note that hydrogen bond formation results in shifts to lower frequencies (longer wavelengths), whereas hydrogen bond breakage results in band shifts to higher frequencies (shorter wavelengths). At the extremes, shifts of the magnitude 80 to 120 cm^{-1} , corresponding to 21 to 50 nm, may be observed. This substantial sensitivity of NIRS to small changes in stereochemistry and hydrogen bonding are part of the reason that moist or reactive synthetic or artificial sample mixtures are often not useful to develop NIR calibrations, particularly if the sample will undergo chemical changes (hydrogen bonding in particular) within a mixture.

17.2.5.2 Particle Size Effects in Forage Materials

Particle size effects are generally multiplicative in nature, as shown in Figure 17.5. It has been deemed desirable to try to eliminate or reduce particle size effects using a variety of mathematical approaches. Multiple linear regression (MLR) will tend to compensate for multiplicative scatter by “forcing” the regression intercept toward the mean reference value of the dataset. This relationship has been demonstrated algebraically [28] and is explained in the chapter on data analysis. Techniques such as derivatizing the data have also been used to minimize the effects of particle size, although derivatives only remove general baseline shifts and cannot correct for multiplicative effects.

More complex means of attacking the particle size question have also been attempted. These include experiments using mathematical modeling for simultaneous removal of particle size and water [29], the use of Fourier deconvolution [30], multiplicative scatter corrections [31], and principal components elimination [32]. Barnes et al. [33] introduced a procedure termed *detrending* that uses standard normal variate (SNV) with polynomial baseline correction [34]. These corrections for particle size may not always improve accuracy of NIRS analysis for two reasons. First, none of these procedures does a perfect job of removing particle size effects independent of absorption information. Second, particle size may be useful information in the calibration even though linear mathematics is used to derive the analytical equation.

These correction techniques cannot be expected to improve calibration and prediction data in all cases. The lack of significant improvement in analytical data using these correction techniques might indicate that the particle size is a potentially useful variable in predicting sample chemistry. The mean particle size and distribution can be an indirect measure of plant chemistry due to the interaction of the various plant structures, such as stems and leaves, to various drying and grinding regimes.

17.3 REFERENCE METHODS FOR NIRS ANALYSIS

17.3.1 DEFINITION OF THE CONCEPT

The NIR spectrum contains information about the major X–H chemical bonds in an agricultural product. The spectrum by definition is dependent on all the functional groups that absorb NIR radiation, which in turn are correlated to the major chemical, physical, and/or sensory components of a substance. The spectrum also contains all of the information due to light interaction with the sample as well as instrumental artifacts, and data collection and computational errors.

In contrast, the currently accepted laboratory procedures that are used to calibrate NIR instruments are not well defined chemically and can be very difficult to relate to spectroscopic data. The problem lies in the fact that the traditional laboratory procedures are the accepted means for measuring the quality of agricultural products and have been in use by the agricultural industry for many decades.

These traditional chemical procedures were developed for one main purpose, that being to estimate the feeding value of a material for livestock production. These procedures fit into three general

categories: (a) proximate analysis, developed more than 100 years ago to estimate general nutritive value; (b) the Van Soest detergent techniques, developed in the late 1960s to give a broad analytical system with better nutritional and chemical definitions; and (c) specific analytical procedures designed to measure single chemical entities that could be used to better describe the nutritional profile of a feed substance.

When NIRS became a viable technology in the 1970s, early researchers were faced with a major dilemma. The agricultural industry was familiar with current laboratory procedures (i.e., reference methods) because it had worked with them for many years and had developed feed composition tables and ration-balancing schemes around these laboratory methods. The response of the agricultural community to changing technology was the requirement that any new analytical procedure must be shown capable of duplicating the analytical results demonstrated by the accepted traditional methods. The difficulty, then, for those working with NIRS on a research basis was that traditional methods were difficult to relate to molecular level chemical and spectroscopic theory. Thus it became a difficult task to obtain forage quality information from spectroscopic data so that livestock rations could be balanced. The difficulties in introducing the technology were compounded by the fact that commercial instruments and software were not developed to the point that industry could rely on their performance.

In the 1980s commercial instruments and software were no longer a limiting factor to the use of NIRS technology. Nevertheless, the basic differences between the traditional laboratory reference methods and the molecular vibrational information available from NIR spectral data are still unresolved.

Two examples will be used to demonstrate this most difficult problem. The most widely used reference method for measuring moisture in agriculture products is oven drying overnight at 100 or 135°C for 2 h. Careful examination of the NIR spectra of these apparently dry samples reveals that moisture is still present in a sample following the 100°C drying method. Both reference drying methods also remove constituents other than moisture in samples containing volatile compounds. Assuming the NIR spectral data are valid, some means of quantifying the differences in the moisture content found using the drying procedures and the apparent moisture content of the sample using a spectroscopic method must be found. Definitions of these relationships are necessary if the agricultural industry is to be able to utilize NIR data. Studies by Windham et al. [35] show that NIRS correlates better to the Karl Fisher method of moisture determination than to oven-drying methods for several types of agricultural products. Thus it has been suggested that oven-drying methods constitute a procedure for estimating a “crude” moisture content, whereas a more valid method of moisture determination is found in the Karl Fisher procedure.

The second example of the difficulty in relating traditional methods to spectroscopic data involves estimating protein content in agricultural products. The current reference method for estimating protein concentration of most agricultural products can be defined as nitrogen times a constant to convert nitrogen to protein. Stated simply, then, crude protein is only a measure of total nitrogen in a material. As was covered in the section on theoretical aspects, nitrogen does not have vibrational response in NIR. By definition, then, the reference method for estimating protein cannot be measured directly by NIRS. NIRS, however, can measure N—H molecular vibrations that are part of the overall protein molecule. It thus should be theoretically possible to estimate protein content, or at least N—H contents, with NIRS. In this case, as in the example of determining moisture, the reference method and NIR measure different attributes of the protein molecule. The strength of the relationship between the reference method and NIRS depend on the agreement between all available N—H bond information and total N.

The relationship between the total nitrogen content of a sample and the total protein predicted using the NIR data is most applicable to those samples containing one major protein in a rather simple matrix. The prediction of crude protein in spring wheat is an example of this case, although the simple relationship becomes complicated when the five major types of wheat are involved. The N—H information becomes more complex with the addition of other proteins and matrix constituents,

and will result in higher prediction errors. Using more wavelength in the prediction equation can compensate to some extent for increased sample complexity but even this approach has limitations. Increasing the number of term in an equation brings about an increase in the effects of instrumental variation such as noise, on the predicted values. In calibration, a mathematical algorithm is used to fit multivariate spectral data containing X—H information to a reference method that is measuring only total reduced nitrogen from amines and amide following Kjeldahl digestion of proteins. Here, then, is a prime example of the dilemma faced by NIR researchers attempting to provide the agricultural industry with an estimate of protein that does not relate well to the current reference method. Predicting animal response using spectroscopic data of forages and grains might be considered to be the greatest challenge to researchers in this field. This application is also considered to be the most valuable potential use of NIRS.

17.3.2 CHEMICAL REFERENCE METHODS FOR AGRICULTURAL PRODUCTS

17.3.2.1 Moisture

The measurement of moisture by NIRS might seem to be relatively uncomplicated due to the high relative intensity or high extinction coefficient of the moisture O—H bands. The O—H vibrations of water exhibit a large absorption in the NIR region and should be easily and accurately quantifiable. Studies by Windham revealed that the accuracies for water determinations were not as great as might have been expected. As previously stated, oven moisture methods are more appropriately termed oven “drying methods” because volatile materials are known to be lost depending on the type of sample. The NIRS calibration error will increase as the concentration of volatile compounds in the sample increases, and as the ratios of pure water to bound or trapped water vary.

17.3.2.2 Protein

Of the many methods used to measure the total nitrogen in a sample of plant material, the Kjeldahl nitrogen determination method is by far the best known and most widely practiced [36,37]. The Kjeldahl procedure was first used in routine agricultural work in 1900 [38]. The actual procedure suffers from several disadvantages, though it has become the reference procedure by which all other nitrogen determination methods are compared. The multiple-step procedure is relatively costly, complicated, time consuming, and hazardous. In an attempt to improve on the disadvantages of the Kjeldahl procedure, other methods for determining ammoniacal nitrogen following the traditional digestion and distillation have been employed [39–41].

As stated earlier, the ratios between total nitrogen and total protein content is variable for each individual type of agricultural product. A number of procedures can and possibly should be tried to replace total nitrogen determination methods as an estimate of total protein. Most of the alternative methods, however, are comparatively imprecise and inaccurate. Possible protein determination procedures might include solubility or peptidization of protein and resulting opalescence, as has been used to estimate protein in wheat [42]. Amino acid analysis by fluorescent strip scanning following gel electrophoresis has been a widely used technique [43]. The use of automatic amino acid analyzers following prior separation by NaDoSO₄ (SDS, sodium dodecyl sulfate) gel electrophoresis is receiving topical usage. Pyrolysis of protein-containing materials with detection of N₂ gas is also utilized for protein content estimates [44].

Reaction of protein with organic dyes has been utilized to estimate the protein content of cereal grains. These methods involve disruption of plant cell walls by ultrasound or homogenization, followed by dye-protein reaction, filtration, and color measurement [45]. None of these techniques adequately measures the total protein content of whole-plant material, for the most part, and so the problem of measuring total protein by a reference method that could be used by NIRS or other

TABLE 17.3
Common NIR “Protein” Absorption Wavelengths as Demonstrated
by the Literature

Reference source	Wavelengths (nm)									
	1000	1500	2000	2500	3000					
Wetzel [46]			2055	2180						
Workman [47]		1778	2139	2190						
Shenk [48]										
Forage filters			1940	2100	2180 2336					
IRSCS				2125	2191					
Barton [18]										
Original				2105	2179 2334					
Revised				2094	2168 2330					
IRSCS	1050		2090							
Kaye [6]										
NH general	1025	1510			2900					
NH bands					2884 2930					
1 amines			1975							
NH perturbed	1460	1498								
Shenk [14]			2100	2180						
Coleman [49]			2100–2200							
Shenk [50]										
Alfalfa protein		1501 1733	2058	2168 2309						
(peaks defined)		1702 1930	1981	2349						
Montana State U. [51]	1680		2100	2180						
Winch [52]										
Legume-grass	1680	1940	2100	2180	2230					
Second calibration	1680	1940	2100	2180						
Burdick [53]	1501	1577	2108	2216						
Coelho [54]	1166	1622	1726	2154						

spectrophotometric methods is still unresolved. The Kjeldahl procedure or modifications of the Kjeldahl method are still the reference method of choice for NIRS calibration of crude protein.

Table 17.3 demonstrates the wavelength selection for several protein calibration equations. Several of the general articles in the bibliography following this chapter are useful for identification of the various bands related to protein content. The most important wavelengths for direct measurement of protein by NIR include the carbonyl stretch of the primary amide at 2060 nm (Maillard effect bound protein observed here); 2168- to 2180-nm combination band consisting of N—H bend second overtone; the C—H stretch/C=O stretch combination; and the C=O stretch/N—H in-plane bend/C—N stretch combination bands. The bands at 2050 to 2060 nm indicate N—H stretching vibrations of various types and are also useful. Aromatic containing amino acids/proteins also exhibit the aromatic C—H stretch first overtone in the 1640- to 1680-nm region. In certain cases the N—H stretch first overtone regions are used for protein measurement in the 1500- to 1530-nm region.

17.3.2.3 Fiber

Animals utilize roughage materials by bacterial action. The bacterial digestion allows the cellulose and pentosans in feed materials to be converted to usable organic acids. The chief acid metabolites include acetic, propionic, and butyric. The volatile fatty acids are absorbed through the ruminal

wall and utilized by the animal for energy. The total cell wall content (or total fiber) of plants is generally estimated using the neutral detergent fiber assay [55]. The total fiber content of forage material includes cellulose, hemicellulose, lignocellulose, silica, pectins, insoluble starch, bound cell wall protein (Maillard reaction), and exogenous contaminants. Plant quality as defined by total digestibility can be estimated by crude fiber or acid detergent fiber assay. The acid detergent fiber method actually defined the total lignocellulose and silica in plant matter. The major portions of a plant cell as delineated by the Van Soest analysis include cell contents and cell wall components.

NIR instruments have been calibrated for the major fibrous components with some success [56]. The structures of cellulose, pectins, and lignin are known to some extent (Table 17.4 and Table 17.5) and thus tentative band assignments can be made for these structures. The cell wall is mostly composed of polysaccharides such as cellulose. The cellulose of plants consists of an unbranched polymer of gluco-pyranose (polyglucose) units linked by carbon (β , 1–4) bonds. The repeating cellulose unit has been termed cellobiose. Cellulose is one of the most abundant, if not the most abundant, carbohydrates on earth. Carbohydrates are of course composed of carbon, hydrogen, and oxygen. Thus we may expect to see many absorption bands in the NIR region for fibrous materials.

Noncellulose polysaccharides are referred to as hemicelluloses: whereas the cellulose gives rise to the various structural microfibrils in the cell wall, the hemicelluloses surrounding the microfibrils compose the matrix cell wall component. One may expect to find several monosaccharide building blocks in the hemicelluloses including glucose, galactose, manose, xylose, and arabinose. Work has been performed to determine the cell wall carbohydrates and starch in alfalfa using NIR [57].

TABLE 17.4
Common NIR “Fiber” Absorption Wavelengths as Demonstrated
by the Literature

Reference source	Wavelengths (nm)									
	1000	1500	2000	2500	3000					
Wetzel [46]			2100	2336	2352					
Workman [47]		1778		2348						
Shenk [48]										
Forage filters		1672	1700	2100	2336					
IRSCS	1403	1679	1708	2374						
Barton [18]										
Original		1539	1661							
Revised		1555	1702							
IRSCS		1615	1710							
Kaye [6]										
CH groups		1700		2350						
Coleman [49]										
Lignin	1100	1300								
Fiber 1		1500								
Fiber 2			2200							
CHO (carbohydrates)				2350						
Shenk [50]				2160						
Cellulose		1483	1930	2102						
Montana State U. [51]		1778	1818	2139	2336					
Winch [52]	1680	1940	2100	2180	2230	2310				
Burdick [53]				2243	2308					
Coelho [54]	1534	1702	2198	2270						
Shenk [15]	1350	1702	1902							

TABLE 17.5
Tentative Band Assignments for Cellulose and Lignin

Type of material	Test band assignment	Wavelength (nm)
Cellulose	O—H str. first overtone	1490
	C—H str. first overtone	1780
	O—H str. and second over C—O str.	1820
	C—H str. and C—H deform	2335
	CH ₂ sym. str. and =CH ₂ deform	2347
	CH ₂ deform. second overtone	2352
	CH str./C—C str. comb.	2488
Lignin	C—H str. second overtone	1170
	O—H str. first overtone	1410
	C—H str. combination	1417
	O—H str. first overtone	1420
	C—H str. combination	1440
	C—H str. first overtone	1685

The vascular portions of leaves and stems are encapsulated by a material called lignin. Late-harvested grasses along with woody plants are relatively high in lignin. Lignin is defined as a member of a group of vegetable compounds that are insoluble in 72% sulfuric acid; these compounds contain phenylpropanoid residues as building blocks.

Other plant materials such as lipids, simple sugars, hormones, vitamins, organic acids, and exogenous compounds are observed in the NIR spectra of fresh plant material. Some of these compounds are lost or altered following drying. This consideration must be noted when working with plant material for either research or routine NIR measurements.

17.3.2.4 Minerals

The concentration of inorganic components in forage crops varies according to crop maturity, temperature, and soil pH and composition. The analyses of mineral content can reveal soil or management deficiencies as well as optimum harvest time for proper crop management. Actual mineral analyses are used to determine the amount of mineral supplementation to be added to an animal ration for proper nutritional balance. Reference methods of analysis include inductively coupled argon plasma (ICP), atomic absorption spectroscopy (AAS), and x-ray fluorescence spectroscopy (XRF). These techniques are well established for the analysis of mineral elements in whole-plant material. The exact procedures for sample preparation and analysis are well documented. Copies of the procedures may be obtained from instrument manufacturers or are readily found using basic texts for each analytical technique.

The major inorganic elements required for the proper growth of forage crop include phosphorus, potassium, sulfur, magnesium, and calcium. The micro nutrients, or nutrients present in trace amounts, are iron, manganese, boron, zinc, copper, molybdenum, and chlorine. Sodium and selenium have most recently received more widespread attention as potential elements measurable using NIRS.

Minerals exist in the plant as organic complexes, chelates with other mineral salts, and in ionic forms. Moreover, the percent dry matter composition of forage material is relatively low in mineral concentration. Because of their low concentrations, the absorptions due to the presence of minerals are difficult to detect above the changes in signal due to the repack errors of the repetitive sample presentation (packing methods). Because the range of mineral composition is low within respective species groups, and even for widely diverse forage sample population groups, it is difficult to select the proper calibration equations based on the computed statistics.

TABLE 17.6

Comparison of Mineral Predictions from a Forage Dataset Using (1) the NIR Spectra with a Four-Term — $\log(1/R)$ equation and (2) Crude Protein and Acid Detergent Fiber Only, to Predict Mineral Content (Mixed Species Hay)

Constituent (%)	<i>N</i>	<i>R</i> ^a	<i>R</i> ^b	SEE ^a	SEE ^b
Calibration					
P	25	0.873	0.840	0.031	0.033
K	25	0.836	0.546	0.347	0.506
Ca	25	0.939	0.892	0.148	0.185
Mg	25	0.873	0.683	0.031	0.045
Prediction					
P	20	0.920	0.910	0.031	0.033
K	20	0.640	0.500	0.495	0.558
Ca	20	0.861	0.857	0.171	0.174
Mg	20	0.358	0.308	0.068	0.070

^a Using four-term — $\log(1/R)$ equation.

^b Using CP and ADF to predict mineral content (% mineral = $a + bx + cy$ where x is the crude protein reference value and y is the acid detergent fiber reference value).

Theoretically, there are no absorption bands for mineral species [58] in the NIR region. Organic complexes and chelates may be detected using NIR, but ionic forms and salts have no spectral fingerprint. Theoretically, salts can be detected in high-moisture samples due to changes in hydrogen bonding are resultant band shifts.

Specific work to determine mineral content in forages indicates that NIR analysis could be directly measuring organic acids [58,59], although there is no special work completed which verifies this hypothesis. On the other hand, it is quite possible that although there is some signal in the NIR region that is directly measuring organically bound mineral components, it is likely that the computed selected calibration wavelengths are related to the protein, fiber, and specular characteristics of the sample. Table 17.6 indicates the estimated vs. actual mineral composition of a forage dataset using only protein and fiber (ADF) information predict mineral composition. More work is required to determine the relationship between the relatively high organic acid composition of forage crops and the individual mineral concentrations.

17.3.2.5 *In Vitro* and Animal Response

Norris et al. [12] described the first prediction of animal response from the spectra of forage samples. Since that time animal studies have been conducted by Wan et al. [60] using esophageal-fistulated sheep and Eckman et al. [61] using sheep fed diets containing mixtures of forage and concentrates. Further response work has been documented by Holecheck et al. [62] using esophageally fistulated cattle, Shenk et al. [63] using sheep fed a hay diet, and Abrams et al. [64] using dairy cattle data. In nearly every case the results of these studies showed that more than one reference method must be used to analyze a diet for best prediction of animal response from that diet. NIRS prediction is best made by using many different wavelengths in a prediction equation that characterizes those spectral bands most correlated directly to animal response for any given diet. The results from each of these studies confirm that given accurate animal response data, the NIRS scanning instrument can accurately predict the intake and digestion of livestock from forage measurement data. Abrams et al. [64]

showed that NIRS was found in some cases to be more accurate in predicting animal response than any of the current reference methods or combinations of these reference methods. This finding again points out the dilemma in relating traditional methods to NIRS. The NIR spectroscopic method can be more accurate in predicting animal response than current reference methods, but it requires calibration against some primary reference procedure to make the predicted information useful for interpretation.

17.3.2.6 Conclusion

Three approaches can be outlined to resolve the differences between components measured using a traditional reference method and the X–H information provided in NIR spectra. First, chemical reference procedures that relate more closely to the NIRS spectroscopic measurement must be selected. Second, spectroscopic reference methods such as nuclear magnetic resonance (NMR) and/or mass spectra [65] should be used to relate directly to NIRS. Third, animal response data should be directly correlated to NIRS data. In relating NIR or other spectrophotometric data directly to animal response, other factors might be incorporated into multi-variate prediction animal response models. These factors include information on animal genetics, health data, weather information, and the like.

17.3.3 SAMPLE PREPARATION METHODS FOR NIRS ANALYSES

17.3.3.1 Sampling

Proper sampling is essential for any chemical technique in order to achieve an aliquot for analysis that properly represents the composition of the larger sample of interest. Increased accuracy and precision in the laboratory will not improve on poor sampling technique, nor will it give more accurate numbers for estimating actual forage or feed quality. Correct sampling procedures are included in the following to serve as a guide to good sampling techniques.

Collect 0.5-kg samples so each forage type to be analyzed. This quantity is considered to be a minimum sample weight for reliable sample representation. For various types of hay, a minimum of 10 to 20 bales or sample sites within the stacks are sampled. A sampling probe is used to collect the several core samples. The core samples are combined, mixed thoroughly, and reduced to 1 L of volume. For ensiled forage materials, each silo or storage site is sampled separately and samples are collected from a minimum of three to five feedings when using an animal management program. Samples must be refrigerated until combined and mixed in a large container. The combined samples are to be frozen before mailing to a laboratory or if analysis will be delayed by 24 h or more. Approximately 1 L of sample is subsampled from the combined larger sample. Haylage, silage, and other high-moisture samples are kept frozen until several hours prior to analyses.

Grains and concentrates are collected as a series from three to five feedings or from several (10 to 15) sampling sites within a storage bin. The samples are then combined, mixed, and a 1-L quantity is withdrawn for analysis.

Pasture is sampled at 15 different locations including the areas of best and worst growth. The material is combined and mixed and subdivided into a 1-L manageable sample.

17.3.3.2 Drying

Two basic methods have been utilized for drying forage samples: oven and microwave. The two methods produce spectra with somewhat different characteristics and thus it is the preferred procedure to use only one drying method for any calibration database. The oven-drying procedure producing the minimum chemical changes in the sample involves a forced air convection oven at 65°C. Samples are left in the oven until they attain approximately 95% dry matter (8 to 12 h).

For microwave drying, a variable-power (not variable intensity) microwave is recommended. Any of the major name brand systems are adequate as long as there is a control for the time interval on

the microwave generator. An internal rack may be purchased to accommodate several paper holders while drying two or more samples simultaneously.

A procedure for microwave drying includes (a) thoroughly mixing the sample; (b) weighing enough sample so as to produce approximately 30 g of dried material; (c) placing the sample in a microwave-approved container; and (d) drying the sample using proper technique as explained in the following text.

Proper microwave drying technique involves using a power setting, and specified time interval for drying, dependent on the oven manufacturer and model number, and the total moisture content within the sample. The important aspects of microwave drying are as follows: (a) To prevent hot spots caused by the intense energy beam characteristic of some microwave generators, the sample should be taken out of the oven and mixed periodically during the drying process. (b) Samples with greater than about 70% dry matter should be mixed every $\frac{1}{2}$ to 1 min (try the shorter time until mastering the technique). (c) Very moist samples, with less than about 65–70% dry matter, can be left in the microwave for 3 to 4 min, then mixed, then dried for additional 30 s intervals. Mixing facilitates the drying process and prevents burning of the samples. Samples may be readily ignited by high internal temperatures as the drying process nears the 90 to 95% dry matter range. Overheating will chemically and thus spectrally affect samples and should thus be avoided.

17.3.3.3 Grinding

Dried samples should be chopped in a commercial grade kitchen-type blender for 30 s to 1 min; then the sample is ground through any one of a variety of grinders. The grinders most commonly used in forage work include the UDY, the Cyclotec, the Wiley, or the Brinkman; all with 0.5- or 1.0-mm screens. The 1.0-mm screen is most commonly used due to the combined advantages of reasonable sample throughput and optimum particle size.

Samples that are to be analyzed in slurry form by transmission or diffuse reflectance may be pulverized or homogenized using various cell disrupter devices or high-speed blenders. The solid particulate material is suspended by vigorous mixing shaking and spectral measurements are taken immediately.

17.3.3.4 Mixing

Samples that are in a dry form require tumbling or thorough mixing to minimize the effects of repack error. Repack error is an effect that is demonstrated for a given sample as the differences between sample spectral measurements of several aliquots of a sample. If this effect is not minimized, the correlation between the optical data and the reference chemical information is drastically reduced and the sensitivity of NIR for any particular application is reduced. Repack averaging, or compression, and other methods of reducing these effects have been demonstrated in the literature [66].

17.3.3.5 Packing

As discussed in the section on light interaction and particle size in this chapter (Figure 17.2), it is essential that the NIR user present samples to the spectrophotometer in the most repetitive or reproducible manner. Table 17.7 delineates the types of presentation devices currently available by the various instrument manufacturers for sample presentations. The variety of techniques available allow several approaches to the forage quality analysis problem. Diffuse reflectance of dried and ground samples in a closed-cup device is the most commonly used procedure, whereas transreflectance and transmission have also been used to analyze wet forage samples [67,68]. Experiments using remote fiberoptic probes have been attempted using the overtone region of the NIR spectrum in forage analyses [69]. Reasonably priced glass fibers attenuate energy, particularly in the combination band region (1.8 to 2.7 μm) due to the relatively high O—H content of the glass. Further improvements in fiber technology will allow greater uses for glass and other materials as fiberoptic components for quantitative as well as qualitative analytical work.

TABLE 17.7
Types of Sample Presentation Devices
Available from Commercial NIR Instrument
Manufacturers^a

- Solid sample cup
- Rotating solid sample cup
- Liquid cup
- Viscous liquid cup
- Transmission liquid cell
- Disposable sample cup
- Transflectance cell
- Research transflectance cell
- Chemical cell
- Thermostatted transflectance liquid cell
- Transmission microcell
- Remote fiberoptics assemblies
- Film holding cell
- Long- or short-pathlength flow cells
- Semimicrotransmission cells
- Low-temperature (77 K) devices
- Various coarse sample cells

^a These devices are sold with various pathlengths and differing mechanical structure depending upon manufacturer. Several manufacturers will custom design cells or cell holders for individual customers.

17.3.3.6 Unground Samples

The trend in NIRS analysis since the 1990s has been to analyze agriculture samples unground and even undried. Grains were analyzed unground in transmission since the late 1980s. Good reference papers for early work with forages are by Dardenne et al. and de la Roza et al. (1995). Samples of these materials are scanned in their natural form and then dried and ground for reference method analysis. If the sample is high in moisture, the dry matter determination must be made from the wet material immediately after it is scanned. The moisture of a sample can change 0.5% in less than a minute so this analysis must be done with great care. In addition, the samples are sometimes not homogenous. Unground corn silage is a good example. To obtain a good representative sample, it is best to use a large sample and repack the cup two to three times to average out sample variation and chemical composition.

In general it can be said that unground samples of grains will provide the same accuracy as ground samples. With high-moisture forage materials, the accuracy is not as good as it is for dried and ground samples. Nevertheless, if the calibration is done carefully, these prediction models are useful in routine analysis.

17.4 CALIBRATING NIR INSTRUMENTATION

17.4.1 COMPILING A SAMPLE/SPECTRAL LIBRARY

In order to develop an analytical method using NIR spectroscopy, a calibration set is used. This calibration set allows the user to collect spectra from a set of samples with widely varying compositions and matrix conditions. These data are collected at several discrete wavelengths where they are processed using a variety of computerized mathematical modeling techniques [99–110]. When

the collected or measured spectral information is regressed against the known concentration values in the calibration set, a series of weighing (regression) coefficients can be determined. When collected spectral information from unknown samples is multiplied by the regression coefficients (stored in the computer), an analytical result is calculated. Both wavelengths and coefficients for the calibration equation are chosen using multiple linear regression mathematics, irrespective of the data decomposition technique(s) used prior to the regression. Data decomposition techniques are mathematical approaches that reduce data in the full-wavelength spectral domain to those of a reduced dataset or series of spectra functions. Principal components analysis (PCA), partial least squares (PLS), derivatives, Fourier space, and even a set of fixed filter wavelengths are all data decomposition techniques. The specifics of these techniques are addressed in several areas of this text.

Selecting a forage sample set for proper calibration has been compared to using pollsters to predict elections: if the sample is large enough and representative of the entire voting population; it will be accurate [70]. Thus the reference laboratory must collect and analyze in duplicate a sample population that is representative of the entire population being analyzed. Calibration sets must not only uniformly cover an entire constituent range, they must also be composed of a uniformly distributed set of sample types. Ideal calibration sets are composed of a proper number of samples with widely varying compositions and constituent values. It was originally suspected, based on the wheat and oilseed data, that a minimum of 50 samples would be required to derive a calibration equation, across forage samples of mixed species [15]. These calibrations were successful for very small populations but did not predict large randomly selected populations adequately. For example, it is now known that to adequately measure constituents across all types of wheat, the screening of hundreds of samples is needed for a good robust calibration. The calibration samples used must contain all of the variance expected to be encountered in a “real-world” analysis situation. The variance of the sample set used to produce NIR calibration equations determines both the robustness and the accuracy of the analytical equation.

17.4.2 REPEATABILITY FILE

A general rule for developing a calibration database is to include all the sources of variation that you expect to encounter during routine analysis. Often, two sources of variation are not well represented in the Product Library file (nomenclature used by NIR Systems). One is variation in temperature, both of the sample and of the instrument. Temperature affects OH bonding, causing a shift in the composite absorption peak for water. This shift is affected by starch content. The other source of variation is instrument differences that remain after instrument standardization. This is less important if several standardized instruments were used to scan samples for the Product Library. Variation sources such as these can be incorporated into the calibration after the database is developed by creating a repeatability (REP) file. A REP file contains spectra of one or more sealed samples scanned under different conditions.

The goal of including a REP file in a calibration is to develop an equation that gives the same predicted value across all conditions represented in the scans. The equation is developed to be insensitive to the REP file’s spectral variation. It is important that there be no changes in the sample scanned under the different conditions. If the sample is allowed to change in moisture content, for example, the equation will be less sensitive to moisture changes. A REP file should not be used to make a calibration less sensitive to repack errors because the variation among repacks is too similar to the variation among samples to be analyzed. The REP file concept is a special feature that is only implemented in WinISI II (a product by Infrasoft International, LLC).

How is a REP file used in the calibration process? First, the mean of all spectra obtained from one physical sample is computed. The mean is then subtracted from each spectrum in the group. This is repeated for each sample used in the REP file, giving a file of spectral deviations. The spectra in the Product Library are also centered by subtracting the grand mean from each individual spectrum.

The calibration program then combines the Product Library spectral deviations with the REP file spectral deviations. The calibration program scales the REP file deviations to be appropriate for the number of Product Library spectra. The centered laboratory reference values from the Product Library are combined with zeros for the REP file. A normal least squares calibration has the property that the predicted value for the average spectrum is the average lab value. A calibration using a REP file attempts to make the predicted value of the average spectrum plus each REP file deviation still equal the average lab value. REP file calibrations are slightly less accurate than their least squares counterparts, but offer much improved repeatability in situations represented in the REP file. REP files are highly recommended.

17.4.3 GLOBAL CALIBRATIONS

A global calibration is one designed to analyze all reasonable samples of a given product. More precisely, a global calibration might have the capability of being used for 90 to 95% of all samples of a given product. The first step in creating a global calibration is to define the domain of samples to be covered by the calibration [71,72]. Specific calibrations based on a small range of samples will typically perform better than broad-based calibrations as long as the samples to be analyzed are represented in the calibration set. However, only a small set of samples can be analyzed by a specific calibration. The goal in deriving a global calibration is to cover as broad a range of samples as possible while maintaining acceptable accuracy [64,73].

17.4.3.1 Sample Selection

Once the analysis domain is defined, samples must be selected to represent the domain. Random sampling could be used if all samples in the domain were available at one time. Global calibrations, however, are usually developed to analyze new samples collected in different locations and in future harvests or years. The best way to prepare for these future samples is to use samples collected from many different locations from recent harvests or years. Several techniques for reducing large forage sample populations to manageable subsets have been attempted. To date these procedures have included selection of sample subsets by means of random, stratified, or nearest centroid clustering techniques [71,72,74,75]. Unique sample selection has been useful as a technique for subset selection. This method provides a subset including only those samples that are most spectrally unique and will provide the maximum sample variance in the smallest number of samples [76].

Two statistics are useful in selecting samples of a global calibration. The standardized H statistic (Mahalanobis distance) tells how different a sample is from the average sample in the calibration set. We will call this the Global H (GH). Adding legitimate samples with large GH values to the calibration set will expand the analysis domain. However, as the domain is enlarged, the NIR spectrum to sample chemistry relationship may become nonlinear, resulting in less accurate models based on linear methods. A good limit for GH is 3.0.

The second statistic is the Neighborhood H (NH) and is used to control the closeness of neighboring samples within the calibration set. The NH of a sample is similar to the GH, but instead of being the distance to the average spectrum, it is the distance to the closest neighboring sample. The neighborhood of a sample is defined as the space around the sample where another spectrum would have a distance to the sample less than a specified limit value. A good limit for NH is 0.60 if scanning from 1100 to 2498 nm or 0.20 if scanning from 850 to 1050 nm. NH values can be used to eliminate nearly redundant spectra by letting one sample represent each neighborhood and eliminating any other spectra in the neighborhood.

Proper use of these statistical tools will provide two benefits: first, the calibration set will be efficiently structured to cover the entire analysis domain, and second, the cost of developing the calibration will be reduced since fewer samples will be required to represent the product.

17.4.4 LOCAL CALIBRATION

LOCAL (Patent # 5,798,526) (added Reference 1 and Reference 2) was developed to predict the one at a time analysis of unknown samples using large databases containing thousands of spectra and reference values. This single sample prediction concept uses the spectrum of the unknown sample to select similar samples to make a custom calibration for each constituent to be predicted. The database is searched for the spectra that correlate most highly with that of the unknown sample. These samples are used to generate a series of PLS models, corresponding to a models with 1, 2, 3, . . . PLS factors. At this point, the global calibration program would pick one model based on the cross validation error. But LOCAL has more information than the global calibration program. LOCAL knows how well the spectrum of the unknown sample is fit by each model. Models that explain the spectrum of the unknown sample better are deemed more trustworthy than models that explain the spectrum of the unknown sample poorly. LOCAL uses this information along with information about the size of the regression coefficients to compute a weighted average of the series of predicted values, giving higher weight to the models deemed most trustworthy. The GH and NH values associated with the prediction are based on PLS scores for each constituent.

In order to provide the highest level of accuracy and analysis speed, the program was designed to work with two different spectra file types. (a) Normal spectra data files in reflectance or transmission, and (b) reduced (RED) derivatized files. Speed can also be improved by splitting a Product Library into subgroups and using a discriminant equation to select the appropriate subgroup database. There is no one best solution for all files of spectra, so testing and experimentation are required to find the combination of data treatments that give the desired accuracy at the fastest speed for a given database. Once the parameters are optimized, the instrument operation program can perform these predictions in real-time. LOCAL can improve the accuracy of prediction over global calibrations by 10% to as much as 30% depending on the spectral diversity of the samples in the Product Library file.

17.4.4.1 Samples with Spectral Features

Global calibration sets must also contain samples with special features. These samples might be the result of an unusual growing season, differing harvesting techniques, contamination, or mixture with other products. Examples of these types of samples include frost-damaged grains, improperly fermented silages early in the harvest season, grains harvested in an immature stage of growth, forage samples contaminated with soil, and mixed feeds derived from unusual combinations of ingredients. Samples with these features may not appear in the normal sample collection procedure.

Spectra may also acquire special characteristics during sample preparation. The best approach is to standardized sample processing and presentation in the instrument. However, small differences in grinding procedures, drying methods, and sample handling inevitably will occur. The calibration set must include samples with these special features. If it is desirable that a calibration support multiple processing methods, samples representing all methods must be added to the calibration set.

Special consideration should be given to sample moisture and temperature as the sample is measured. Unless samples are stored in sealed, moisture-proof pouches, water and other volatile compounds will equilibrate to the levels in the surrounding atmosphere. Samples stored for a long time will tend to equilibrate to a common dry matter value. Samples whose spectra exhibit low and high water content should deliberately be added to the calibration set.

Temperature has a major effect on the shape of water bands in agricultural materials. The underlying cause behind the changes in band shape is explained by the fact that the apparent water band is actually a composite of several underlying bands, representing the different forms of water in the sample. As temperature changes, water in the sample changes from one form to another (i.e., there are different amounts of hydrogen bonding), resulting in a different composite band. A recommended temperature for samples during measurement should be established, and an effort should be made to

deliberately add temperature variability to the scans of calibration samples. This will allow accurate analyses of samples that deviate from the recommended temperature.

In general, deliberately adding samples with special characteristics to broaden the calibration set is beneficial. The advantage of using a calibration specific to a group of samples is that the analyses are usually more accurate than analyses from a broad-based calibration. The disadvantage of using a specific calibration is that many different equations are needed to analyze all possible samples of a product.

17.4.5 MATHEMATICAL TRANSFORMATIONS OF SPECTRAL DATA

As discussed in the chapter on data analysis, there are many techniques that can be used for mathematical modeling of forage calibration data. The proper use of these techniques, however, requires an in-depth knowledge about the statistics relating to repetitive measurements and, specifically, multivariate statistics. Lack of knowledge in these subject areas may result in an improper calibration set, modeled with an empirical-only data-handling technique. Improper interpretation of a calibration equation is a danger when the user is not aware of the meaning of various statistical methods. Simply running data through the latest mathematical algorithm will result in nothing interpretable and is only pseudoscience.

17.4.5.1 Quantification

Several mathematical and statistical procedures are available to calibrate NIRS instruments (i.e., quantify the relationship between spectral data and the reference methods). These methods include Fourier transformation [77], PCA [78], PLS, modified stepwise least squares regression [20], all-possible-pairs regression, least-squares curve fitting [79], and others. Experience in the NIRS Forage Network indicates that least-squares regression can be used to develop global calibrations with acceptable accuracy and performance. Colinearity or intercorrelation do not present themselves to be a problem when using the currently available regression algorithms. For optimal calibrations, the primary requirements include properly measured spectroscopic data combined with accurate and precise reference values. Given these two requirements, modern desktop computers using least-squares algorithms are capable of automatically processing hundreds of samples, testing and evaluating a number of data transformation procedures conducting within-file evaluation, and testing the equation for transferability across instruments.

17.4.5.2 NIRS Algorithms

The purpose of the various algorithms for calibration is to optimize a mathematical interpretation (model) of the calibration data that will allow an accurate prediction of constituent values in unknown samples. The mathematical modeling can be done in a number of ways, but the final test of any calibration is the accurate prediction of samples not included in the calibration. The most rigorous test for a calibration equation is to test its accuracy using a number of sets of known prediction samples and to verify the resulting errors with statistical hypothesis tests. Some of the tests most commonly used in agricultural product calibrations are explained in Section 17.7.

17.4.6 SELECTING CALIBRATION EQUATIONS

Selecting calibration equations for routine or research analysis requires familiarity with the NIR forage literature. Of all the applications of NIR analysis, by far the greatest number of papers to date have been generated in regard to forage and grain analyses [47,80–95]. As seen in the references of this chapter, the main theme of the literature has been the analysis or the development of the

feasibility of using different instruments and math treatments for routine analysis on rather large or potentially open populations.

Statistical criteria alone may be used to select and validate calibration equations in NIR spectroscopy. In the ideal case, much is known regarding the chemistry of the samples to be analyzed. Chemical knowledge allows the user to determine optimal wavelength selection and gives greater confidence that the NIR method is directly predicting the constituent(s) of interest. Without chemical knowledge the user must rely on purely statistical selection criteria and is cautioned against misuse of the computational power routinely available with modern computers.

The method of calibration equation selection developed by the NIRS Forage Network follows in an abbreviated form. Note that the terminology used is that of the Forage Network and it is not consistent with all NIR users or statisticians. Each term and abbreviation will soon be standardized by ASTM Task Group on NIR E-13.03.03. For each constituent, the standard error of calibration (SEC) and coefficient of determination (R^2) for each calibration equation are important criteria for decision making with respect to equation selection. As wavelength (independent terms) are added to the equation, SEC will decrease and R^2 will increase.

A second step in equation selection is to evaluate the standard error of analysis (SEA) of each equation using a sample set not included in the calibration. The SEA is an indication of the performance of the equation on the independent samples. For each math treatment the equation SEA will reach a minimum value as terms are added to the equation. Unlike SEC, which decreases with each additional term, SEA decreases until overfitting of the data is evidenced by an increase in the SEA value as terms are added to the calibration equation. As a guide, the values of SEC and SEA should be within 20% of each other. Other general guidelines for equation selection might include the following. A more detailed chapter on calibration is included elsewhere in this text:

1. The math treatment should have the lowest SEC value and the fewest number of terms (wavelengths) to prevent overfitting.
2. No wavelength should have a partial F -statistic value of less than 10 for its corresponding regression coefficient to prevent overfitting.
3. Regression coefficients should not exceed $\pm 10,000$ in order to minimize problems in equation transfer, due to instrument noise and other inter-instrument variations.
4. The slope of the regression line relating the NIRS analytical value of SEA to the primary reference value should be close to one.

Finally, equation selection is completed using criteria such as those listed in order to choose the final “best” equation from the group of individual best equations. Check the wavelengths for the final selected equation against the literature (such as Table 17.2 to Table 17.5). When the best equation is selected, recalculate a new equation using the same math treatment and wavelengths as had been used in the best equation. This final equation is determined using the original calibration sample data and the prediction or test sample data. This equation is now ready to enter a system to monitor equation performance with time and sample changes.

If the equation performs well at first but its accuracy begins to deteriorate over time, recalibration is probably indicated. Samples should be screened with a subset selection technique and a discriminant test (such as the H statistic), and selected samples should be analyzed with a primary reference method. These selected samples are added to the original calibration set and the equation selection process is repeated.

17.5 ROUTINE AGRICULTURAL APPLICATION

Routine forage analysis involves the setup of a rugged instrument to perform fast and accurate analyses of any forage types received routinely at any analytical laboratory. Routine use of NIR

spectroscopy is a specialized use of the technology requiring substantial problem solving prior to implementation. Routine performance characteristics of NIR are significantly different from research applications. Routine calibrations are generally developed for large groups of plant species with diverse growth stages and thus are expected to include wider variation than research-type calibrations. Routine calibrations are developed using large closed populations in an attempt to mathematically model open populations. Work specifically dealing with the problems of routine analyses are discussed in great detail in Reference 47 and Reference 75 to Reference 94.

Global calibrations are presently available for most of the major products used in formulating diets for livestock. General categories include forage, grains, mixed feeds, and feed ingredients. SEA for these equations falls into three categories. In the first category are those product constituents that exhibit a high correlation between the spectral information and the reference values. Examples include crude protein and moisture in most grains [96]; oil in protein supplements, such as soy-meal; and fat in several types of animal by-products. These product and constituent combinations demonstrate the smallest SEC for any agricultural products in which NIRS is used.

A second category of materials exhibiting slightly larger SEC includes crude protein and moisture in forages and mixed feeds, and in some cases, the fiber and mineral calibrations involving foragelike materials. For these constituents, NIRS and the reference methods are not measuring the same functionality (e.g., NIRS measures N—H, not all forms of N). The higher analysis error in the second category is due to the imperfect relationship between the two measurements. A third category includes all other constituent measurements that have some predictive accuracy that in a specific case are more accurate than average values from a feed composition table.

An example of the latter category is the prediction of crude protein in fish meal. The standard error of analyzing crude protein in fish meal is large when compared to prediction of crude protein in wheat (i.e., 1.5% for fish meal and 0.3% for wheat). This is not a problem of sampling or mathematical method of developing the calibration, but is due to the lack of agreement between the spectral information in the sample and the reference method. The alternatives to accepting the current standard error of 1.5% are to (a) use the average value for crude protein from a feed composition table, (b) wait until the sample is analyzed by the reference method, or (c) calibrate to another reference method that better relates to the information in the spectra. Step (c) also includes educating the industry on the use of this new analysis.

17.6 CALIBRATION TRANSFER

17.6.1 BASIC PROBLEM DEFINED

Calibration development using NIR instrumentation for agricultural products is an involved and tedious undertaking. Therefore it is advantageous to be able to transfer calibrations from the original or “master” instrument to a “slave” or host instrument with minimal loss in performance. This is no minor problem due to the usual differences among instruments (Table 17.8). General guidelines for calibration (or spectral) transfer procedures are described in the following text. As improvements are made in the performance specifications for spectrophotometers, the demand for increased uniformity among instruments increases. The end user is becoming more aware of instrument design specifications and expects nearly identical performance from each instrument used.

17.6.2 METHODS OF TRANSFER

One of the important goals of NIRS analysis has been uniformity of analysis, in otherwords, a sample should obtain the same analysis value from all instruments. The first effort by instrument manufacturers was to accomplish this goal using a slope-and-bias adjustment applied to analysis values. In 1986 a second method was developed that could be used with instruments generating spectra [97], in otherwords, monochromator or tilting filter instruments. This method adjusted the

TABLE 17.8
Factors Affecting Calibration Transfer and Instrument Performance

Factors	Effect
Optics	
Wavelength accuracy	Large
Wavelength reproducibility	Medium
Photometric energy (accuracy)	Small
Dynamic range	Large
Sphere gain	Medium
Specular sensitivity	Medium
Focusing lens characteristics	Medium
Optical glass composition	Small
Optical F ratio	Medium
Stray light effects	Medium
Distance of sample to optics	Large
Resolution	Small
Electronics	
Detector position	Small
Detector photometric response	Medium
Detector linearity	Small
A/D digitization loss	Small
White noise	Small
Response	Small
Temperature stability	Medium
Sample	
Reflectance stability	Medium
Functionality changes with temperature	Large
Specular variability	Medium
Nonhomogeneity	Medium
Calibration	
Wavelength choice	Medium
Coefficient choice	Medium–large
Number of terms	Medium–large
Math treatment	Small

equation wavelengths and coefficients so that the modified equation could be used on different instruments without slope-and-bias adjustment to analytical values.

The third method involves a spectral modification technique. The procedure termed *virtual master*, is performed by developing a virtual template of a master filter or monochromator instrument using a standardization set of samples with widely varying spectral composition. The slave or transfer instrument is then standardized using the virtual template of the master [97].

This concept involves two methodologies: transformation of the master monochromator spectra to look like spectra from another instrument model, and standardization of spectra to look like spectra from the master instrument of the same model. Transformed spectra are said to come from a “virtual” instrument because the instrument from which the spectra appear to come did not produce the spectra. A monochromator can be configured as a virtual tilting or fixed filter instrument. These virtual filter instruments can be used as master instruments to which actual filter instruments are standardized. Both methods rely on the spectra of the standardization samples to characterize the instruments.

Verification of instrument standardization is found in Table 17.9. This table represents the standard error of difference (SED) between the master monochromator maintained in the United States and

TABLE 17.9
SED between Master Monochromator and Three 6250 European Monochromators Using the Same Ring Test Calibration for All 221 Samples^a

	IUK	IB	IG	UKB	UKG	BG
DM	0.98	0.87	0.69	0.77	0.93	0.69
N	0.29	0.35	0.26	0.52	0.35	0.34
DOV	4.10	6.64	5.32	3.59	2.87	3.23
DOC	4.01	5.39	4.43	4.69	1.66	4.74
ASH	0.32	0.28	0.33	0.23	0.19	0.23
CF	0.31	0.25	0.27	0.19	0.22	0.23

^a All measurements are in g/kg. I = Infrasoft International; UK = United Kingdom, Ralph Barnes; B = Belgium, Dr. R. Bistone; G = Germany, Dr. Christian Paul. DM = dry matter. N = nitrogen. DOV = digestible organic matter *in vitro*. DOC = digestible organic matter in cellulase. ASH = ash. CF = crude fiber.

TABLE 17.10
SED between Four Standardized Pacific Scientific 4250 Instruments Using the Same Wheat Analytical Equations^a

Instrument # →	1	2	3	4
Crude protein	0.50	0.40	0.55	0.46
Fiber	3.20	2.25	2.30	3.30
Dry matter	0.70	0.80	0.72	0.88

^a All measurements are in g/kg.

three other monochromators in Europe. Calibration equations for this test were developed from 221 samples supplied for the European NIRS ring test on grass silage. Agreement between monochromator analysis after standardization is almost as good as agreement of blind duplicates in the respective reference method in a good laboratory.

Comparisons of standardized tilting filter instruments in Belgium are presented in Table 17.10 and Table 17.11. Table 17.11 shows the agreement between four Pacific Scientific 4250 standardized instruments analyzing soymeal. Table 17.11 shows the agreement of the same instruments analyzing hay. SED were all acceptable and similar to those reported in USDA Handbook 643[20] for calibration equation transfer. The advantage to these methods is that calibration equations can be shared among all standardized instruments without modification.

A single Technicon 400 was available for this study; it was used to generate a virtual 400 from the master monochromator. The SEDs reported in Table 17.12 represent the accuracy of transforming a Pacific Scientific monochromator into a virtual Technicon 400; they do not include the variability from one Technicon 400 to another.

17.6.3 NEW METHODS OF INSTRUMENT STANDARDIZATION

Following the development of these methods of standardization, it was found that a simpler method of standardization could be employed for the NIR Systems line of instrumentation. Because wavelength alignment is accomplished with internal polystyrene and didymium filters within the instrument,

TABLE 17.11
SED between Four Standardized Pacific Scientific
4250 Instruments Using the Same Hay Analytical
Equations^a

Instrument # →	1	2	3	4
Crude protein	2.10	1.80	2.22	1.99
Acid detergent fiber	6.72	5.99	7.01	6.40
Neutral detergent fiber	10.50	11.10	9.98	10.65
Phosphorus	0.10	0.13	0.15	0.12
Calcium	0.42	0.52	0.44	0.59
Potassium	0.82	0.90	0.72	0.95
Magnesium	0.11	0.13	0.15	0.10
Dry matter	1.30	1.10	1.19	1.50

^a All measurements are in g/kg.

TABLE 17.12
SED between the Master
Pacific 6250 Monochromator
and Standardized Technicon
400 Using the Same Wheat
Analytical Equations^a

Instrument # →	1
Crude protein	0.50
Fiber	2.35
Dry matter	0.86

^a All measurements are in g/kg.

no further wavelength adjustments were necessary. The entire standardization procedure could be simplified to scanning one sealed sample of the product. The best single sample to accomplish this standardization is the sample closest to the mean of the Product Library of samples. By scanning this sealed sample on both master and host instrument, the standardization could be accomplished by simple subtraction of master and host spectra. This simple correction is then made before the unknown sample is predicted in routine analysis.

The standardization procedures were expanded further to instruments measuring transmission through solid particles between 850 and 1050 nm. Under these conditions, three adjustments needed to be made: (a) wavelength alignment needs to be verified, (b) pathlength needs to be adjusted, and (c) spectral offset needs to be adjusted. These corrections were successfully made for Infratec instruments using three samples of grain. Because these samples cannot be sealed, care must be taken to (a) choose three samples with GH values near the center of the Product Library, (b) make sure each sample pours uniformly, and (c) make sure the samples are equilibrated to room moisture. The samples must be scanned in triplicate and averaged to reduce the effect of repack error with poured samples.

17.6.4 CALIBRATION TRANSFER GOAL

The goal of standardization is to have the total error (bias and unexplained error) among instruments of the same magnitude as the variation of predicting subsamples from a container of the product.

In addition, the GH and NH outlier tests should have the same sensitivity on the host instruments as they do on the master instrument. Our experience over the past 20 years with standardization of NIRS instruments, monochromators manufactured by others, filter instruments, and now the Infracore instrument has shown that this goal can be achieved using the methods described in this chapter. They have worked for ground and unground samples, dry and high moisture samples, solids, liquids and mixtures of the two, filter and monochromator instruments, and in reflectance and transmission.

Two methods can be used to provide statistical verification of these results. The first is the SED among predicted values for sealed samples (or room-equilibrated samples for an Infracore) between the master and host instrument.

$$SD = \frac{\text{SQRT}[\text{SUM}(\text{master}_i - \text{host}_i)^2]}{\text{SQRT}[n - 1]}$$

The second method is applied when more than two instruments are to be compared. In this case we use the standard error of pooled or average differences (ASD) among instruments. Some confusion occurs in that the formulas differ by the SQRT of 2.0 when applied to two instruments. The pooled error formula gives a smaller value.

$$ASD = \frac{\text{SQRT}\{\text{SUM}[\text{SUM}(\text{replicate}_{ij} - \text{mean}_i)^2]\}}{\text{SQRT}[\text{SUM}(n - 1)]}$$

In the situation where more than two instruments are to be compared, an analysis of variance can be used to remove sample and instrument differences and compare bias and unexplained error for the group of standardized instruments.

17.6.5 SUMMARY OF MAJOR POINTS

The transformation to a virtual master and instrument standardization procedures allows the user of NIR instrumentation to be in control of the analytical process. Key points to remember are:

1. Transformation and characterization requires an external IBM-compatible computer having 512K RAM memory, 8087 math coprocessor, chip, and EGA card.
2. The NIRS instrument must be in good working condition with RMS noise less than 30 μ OD (ceramic reference to ceramic).
3. Every effort should be made to stabilize instrument and sample temperature as the characterization samples are scanned.
4. Reference samples to be used for characterization should represent all products to be analyzed by the instrument. A minimum of 30 samples is recommended.
5. The transformation of a scanning monochromator to filter instrument need only be developed once for each instrument type.
6. Instrument standardization requires scanning the characterization samples with the master monochromator (as a virtual filter instrument) and the filter instrument to be standardized.
7. Instruments need only be standardized once unless the instrument undergoes a major change or repair.
8. The instrument controller program used in routine operation of the instrument standardizes the spectra for sample analysis and spectral storage.
9. Calibration databases are maintained only on the master monochromator and used to develop calibration equations for filter or other monochromator instruments.
10. Equations for any product or constituent can generally be used among standardized instruments without slope or bias adjustment.
11. Equations for a product or constituent can be updated at any time and made available to host instruments even if new math treatments or wavelengths are selected.

12. Analytical results across instruments are most repeatable when all instruments of the same model use the same equation.
13. A simple monitoring system can be employed to establish and maintain confidence in the entire system.

17.7 MONITORING NIRS ANALYSES

17.7.1 STANDARD ERROR OF A DIFFERENCE

The risk involved in applying broad-based NIRS analytical equations to future samples has been discussed elsewhere [29,64,98]. Many factors can cause the analysis error to exceed the limits established in the calibration process. In order to test for this situation, choose a small set of samples and compare the NIRS analytical and reference method values. Control limits for these errors are established from the SEC for each calibration equation.

The monitoring system is primarily based on the concept of differences. One of the most meaningful statistics used in this monitoring system is the SED. The following terms are used in the monitoring system as it is used by the agriculture industry:

RMS	standard error of difference between two scans collected from the same instrument.
SEC	standard error of calibration or standard error of difference between the reference method values and NIRS analysis values in calibration using degrees of freedom = $n - p - 1$.
SEA, SEP	standard error of analysis, or standard error of performance, expressed as a difference between NIR analyses values and reference method values.
SEL	standard error of difference between blind duplicates using the reference method.
SED	standard error of difference between reference values obtained from different laboratories, standard error of difference between NIRS analysis values between two standardized NIRS instruments of the same model, standard error of difference between NIRS analysis values between two standardized NIRS instruments of different models.

The formulas for these calculations are as follows:

$$SED = \sqrt{\frac{\sum D^2}{N}}$$

$$Bias = \frac{\sum D}{N}$$

$$SED(C) = \sqrt{\frac{\sum D^2 - (\sum D)^2/N}{N - 1}}$$

As shown by the above formulas, SED can be partitioned into two errors: (a) a systematic difference (bias) between the two sets of analysis values calculated as the difference between the means of the two sets of values, and (b) SED corrected for bias (SED(C)) or unexplained error, caused by using an equation not designed specifically for the samples being tested. The slope of reference values regressed on predicted values usually differs from 1, and is included in unexplained error.

17.7.2 CONTROL LIMITS

Having defined the errors, two control limits are needed: one to determine if a meaningful bias is occurring and one to determine if a meaningful increase in unexplained error is occurring. Details of

this procedure are described in a supplement to USDA Handbook 643 published in 1988; however, a brief example of the monitoring system is presented in the following text.

Assuming that (a) a difference between the NIRS analytical values and reference values in either bias or unexplained error is to be detected with 90% confidence, (b) at least 100 samples are present in the calibration set, and (c) a bias greater than the SEC and an unexplained error greater than two times the SEC are unacceptable, the following procedure is recommended:

1. Choose nine samples at random from the group of samples to be analyzed.
2. Obtain analysis values by both the reference and NIRS methods.
3. Calculate bias control limits: $BCL = 0.6 (SEC)$.
4. Calculate unexplained error control limits: $UCL = 1.3 (SEC)$.
5. Compute the bias and unexplained error of the analytical values in the test set.

These tests should be made whenever there is some reason to believe that an equation may not be performing well. At a minimum, at least one sample out of every 100 should be set aside for reference method analysis after NIRS analysis. One in 50 would be better. When nine monitoring samples have been accumulated, they should be analyzed by the reference method and evaluated using the method described.

If bias or unexplained error control limits are exceeded, steps should be taken to either add these samples to the existing calibration set and recalibrate, or develop a new calibration for this specific group of samples. Adjusting the bias only fixes the problem on a temporary basis; another set of nine samples would probably suggest a different bias correction. The problem is in the calibration set and can only be corrected there.

This simple monitoring system using control limits can be applied to other areas of the analytical system. These include errors within laboratories (SEL), errors between laboratories (SED), errors between standardized instruments of the same model (SED), and errors between instruments of different models (SED). General guidelines for the monitoring system are given in Table 17.13.

TABLE 17.13
Control Limits for Monitoring NIRS Technology

Control limits	Value
Control limits for a properly working instrument	$RMS < 30$
Control limits for a properly standardized instrument	$SED < 1.3$ (factory SED)
Control limits for a correct set of test samples	H average < 2.0 , none > 3.0
Control limits for duplicate reference values from another laboratory	$SEL < 1.3$ (M SEL)
Control limits for agreement between NIRS analysis values and reference method values	$SEA < 1.3$ (SEC)
Control limits for agreement of reference method values between two laboratories	$SED < 1.3$ (M SEL)
Control limits for agreement of NIRS analysis values between two standardized instruments of different models	$SED < SEC$

RMS = root mean square noise of the instrument. SEC = standard error of calibration for each constituent. SEA = standard error of analysis or the difference between analysis values from the same samples analyzed by NIR and the reference laboratory. SED = standard error of a difference. Factory SED = standard error of the difference between the same samples analyzed by the master and slave instrument at different times. H = standardized H statistic. SEL = standard error of the laboratory reference values. (This statistic can be either the difference between duplicates in one laboratory or the difference between the same samples analyzed by two different laboratories.) M SEL = standard error of blind duplicates in the master reference laboratory.

17.7.3 ACTION

The action required if any of these control limits are expected depend upon the specific test performed. If an instrument is not working properly or is not standardized properly it usually is a problem to be taken care of by the manufacturer. If a problem exists with the reference methods within or between laboratories, these problems have to be addressed there. If the control limits for the analysis have been exceeded and the problem is not an instrument, standardization, or reference method problem, the calibration itself may be at fault.

The most obvious problem with the calibration may be that the new test samples were not processed in the same manner as the calibration samples. If this is found to be the case, either new samples must be taken and processed correctly or a new calibration must be performed. Only after all of these points have been exhausted should the calibration equation itself be altered. Simple slope and bias adjustment will probably do more harm than good.

Slope and bias changes can be caused by changes in the reference method changes in the instrument, changes in sample processing, and changes in sample temperature. The problem can best be corrected by eliminating all of these problems. If the problem persists, the only way it can be corrected is by adding the new test samples to the calibration. This action will require recalibration.

17.7.4 RECALIBRATION

Recalibration is necessary whenever (a) the instrument changes, (b) samples are believed to be part of the original calibration population but on analysis exhibit many large H values, or (c) during monitoring of NIRS analyses using the primary reference method, the test results continue to fall outside the control limits.

Samples for recalibration should be selected on the basis of spectral characteristics. As was stated in Section 17.4.2, the same program used to select samples for the calibration can now be used to choose samples for recalibration. The current calibration file is used as a library and the new population of samples is divided into similar samples, samples to be added to the calibration, and outlier samples.

The samples to be added are analyzed by the primary reference method and the new population of calibration samples resubmitted to the entire calibration procedure. After the "best" equation for each constituent has been selected, continue to monitor the analysis of this new equation. Depending on the spectral variation of the new population to be predicted, the calibration, monitoring, and recalibration procedure many need to be repeated a number of times.

REFERENCES

1. W. Gordy and P. C. Martin, *J. Chem. Phys.*, 7: 99 (1939).
2. W. Gordy and S. C. Stanford, *J. Chem. Phys.*, 9: 204 (1941).
3. W. Kaye, C. Canon, and R. G. Devaney, *J. Opt. Soc. Am.*, 41: 658 (1951).
4. W. Kaye and R. G. Devaney, *J. Opt. Soc. Am.*, 42: 567 (1952).
5. W. Kaye and A. Thompson, Hydrogen Bonding and Electrostatic Interaction, Molecular Symposium, Ohio State University, June 1953.
6. W. Kaye, *Spectrochim. Acta*, 6: 257 (1954).
7. W. Kaye, *Spectrochim. Acta*, 1: 181 (1955).
8. K. B. Whetzel, *Appl. Spectrosc. Rev.*, 2: 1 (1968).
9. I. Ben-Gera and K. H. Norris, *Isr. J. Agr. Res.*, 18: 125 (1968).
10. I. Ben-Gera and K. H. Norris, *J. Food Sci.*, 33: 64 (1968).
11. K. H. Norris and R. F. Barnes, Infrared reflectance analysis of nutritive value of feedstuffs, in *Proc. First Int. Symp. Feed Comp.*, Utah Agr. Exp. Sta., Utah State University, Logan, 1976, p. 237.
12. K. H. Norris, R. F. Barnes, J. E. Moore, and J. S. Shenk, *J. Anim. Sci.*, 43: 889 (1976).

13. J. S. Shenk and M. R. Hoover, Infrared reflectance spectro-computer design and application, in *Proc. 7th Technicon Intl Cong.*, 2, Tarrytown, NY, 1976, p. 122.
14. J. S. Shenk and R. F. Barnes, Current Status of Infrared Reflectance, in *Proc. 34th South. Past. and Forage Crop Impr. Conf.*, Auburn, Alabama, 1977, pp. 57–62.
15. J. S. Shenk, M. O. Westerhaus, and M. R. Hoover, Infrared reflectance analysis of forages, in *Proc. Int. Grain and Forage Harr. Conf.*, Am. Soc. Agr. Eng., St. Joseph, MI, 1978, pp. 242–244.
16. J. S. Shenk, M. O. Westerhaus, M. R. Hoover, K. M. Mayberry, and H. K. Goering, Predicting Forage Quality by Infrared Reflectance Spectroscopy, in *Proc. 2nd Int. Green Crop Drying Conf.*, University of Saskatchewan, 1978, pp. 292–299.
17. F. E. Barton II and D. Burdick, Analysis of bermudagrass and other forages by near-infrared reflectance, in *Proc. 8th Res. and Ind. Conf.*, Coastal Bermudagrass Processors Assoc., Athens, Georgia, 1978, pp. 45–51.
18. F. E. Barton II and D. Burdick, *J. Agr. Food Chem.*, 27: 1248 (1979).
19. F. E. Barton II and D. Burdick, Prediction of Crude Protein in Dehydrated Coastal Bermudagrass by NIR Reflectance, in *Proc. 10th Res. and Ind. Conf.*, Coastal Bermudagrass Processors Assoc., Athens, Georgia, 1980, pp. 103–106.
20. J. S. Shenk, in *Near Infrared Reflectance Spectroscopy (NIRS); Analysis of Forage Quality*, G. C. Marten, F. E. Barton II, and J. S. Shenk (eds.), USDA Agricultural Handbook No. 643.
21. C. Roberts, J. Workman, and J. Reeves (eds.), *Near-Infrared Spectroscopy in Agriculture*, ASA-CSSA-SSSA, Madison, WI, June 2004.
22. G. C. Marten, J. L. Halgerson, and D. A. Sleper, Near Infrared reflectance spectroscopy evaluation of ruminal fermentation and cellulase digestion of diverse forages, *Crop Sci.*, 28: 163–167 (1988).
23. D. H. Clark and R. C. Lamb, *J. Dairy Sci.*, 69: Abstr. Suppl., 136 (1986).
24. J. Workman, *J. Dairy Sci.*, 69: Abstr. Suppl., 136 (1986).
25. B. G. Osborne and T. Fearn, *Near Infrared Spectroscopy in Food Analysis*, Wiley, New York, 1986, pp. 35–40.
26. J. R. Dyer, *Applications of Absorption Spectroscopy of Organic Compounds*, Prentice-Hall, Englewood Cliffs, NJ, 1965.
27. Bran and Luebbe/Technicon Instruments, *Instruction Manual for IDAS: Theory of NIRA*, 1987, pp. A1.9–A1.11.
28. J. Workman, Near-Infrared Spectroscopy: Evaluating the Effects of Known Spectral Changes on Wavelength Selection and Optimal Mathematical Treatment Using Synthetically Generated Data, *Eastern Analytical Symposium*, New York, 1986, No. 269.
29. J. S. Shenk and M. O. Westerhaus, *Software Standardization of NIR Instrumentation*, 14th FACSS, 1987, No. 129.
30. W. F. McClure and R. E. Williamson, Status of Near Infrared Technology in the Tobacco Industry, in *Proc. 40th Tobacco Chemists Research Conf.*, Knoxville, TN, 1986, pp. 34–35.
31. H. Martens, Multivariate calibration by partial least squares (PLS) regression, *Eastern Analytical Symposium*, New York, 1986, No. 85.
32. D. Honigs and C. Miller, Regular and anomalous particle size effects in near infrared spectroscopy, *Eastern Analytical Symposium*, New York, 1987, No. 135.
33. R. J. Barnes, M. S. Dhanoa, and Susan L. Lister, *Appl. Spectrosc.*, 43: 772–777 (1989).
34. J. S. Shenk and M. O. Westerhaus, New standardization and calibration procedures for NIRS analytical systems, *Crop. Sci.*, 31: 1694–1696 (1990).
35. W. R. Windham, J. R. Robertson, and R. G. Leffler, *Crop Sci.*, 4: 777–783 (1987).
36. J. Kjeldahl, *Z. Anal. Chem.*, 22: 366 (1883).
37. J. Kjeldahl, *Medd. Carlsberg Lab.*, 2: 1 (1883).
38. W. O. Whitcomb and E. J. Bell, The protein test in marketing wheat, *Montana State Coll. Agr. Expt. Sta. Bull.*, 189 (1926).
39. Orion Research, Inc., Applications Bulletin Nos. 16 and 17, Cambridge, Mass., 1978.
40. Association of Official Analytical Chemists (AOAC), *Official Methods of Analysis*, 13th ed. Washington, DC, 1980.
41. J. Wong, *J. Biol. Chem.*, 55: 431 (1923).
42. L. Feinstein and J. R. Hart, *Cereal Chem.*, 36: 191 (1959).

43. F. R. Elevitch, S. B. Aronson, T. V. Feichtmeir, and M. L. Enterline, *Am. J. Clin. Pathol.*, **46**: 692 (1966).
44. H. H. Cheng and J. M. Bremmer, *Methods of Soil Analysis*, Am. Soc. Agron., Madison, WI, 1965.
45. D. C. Udy, *Cereal Chem.*, **31**: 389 (1954).
46. D. L. Wetzel, *Anal. Chem.*, **55**: 1172A (1983).
47. J. J. Workman, doctoral dissertation, Columbia Pacific University, University Michigan films International No. LD-00993, 1984.
48. J. S. Shenk, M. O. Westerhaus, and M. R. Hoover, *J. Dairy Sci.*, **62**: 810 (1979).
49. S. W. Coleman, F. E. Barton II, and R. D. Meyer, Calibration of a Near-Infrared Spectrometer for Prediction of Forage Quality, Oklahoma State University A Exper. Station, June 1982, pp. 104–105.
50. J. S. Shenk, K. H. Norris, R. F. Barnes, and G. W. Fissel, Forage and Feedstuff Analysis with Infrared Reflectance Spectro/Computer System, *XIII Interna Grassl. Cong.*, Leipzig, May 1977, pp. 1440–1441.
51. P. Meland, Ingman Laboratories, Minneapolis, MN, personal communication, 1980.
52. J. E. Winch, Crop Sci. Dept., Univ. of Guelph, Ontario, Canada, personal communication, 1980.
53. D. Burdick, F. E. Barton II, and B. D. Nelson, Rapid Determination of Forage Quality with a Near Infrared Filter Spectrometer, in *Proc. 36th Southern Pasture a Forage Crop Improvement Conf.*, Beltsville, Maryland, May 1979, pp. 81–86.
54. M. Coelho and F. G. Hembry, Laboratory Methods of Forage Evaluation. III. Ne Infrared Reflectance Analysis, in *Proc. of the Program and Research Progress Reports*, Louisiana State University Anim. Sci. Dept., Baton Rouge, 1982, pp. 195–199.
55. H. K. Goering and P. J. Van Soest, *Forage Fiber Analyses*, USDA Agricultural Handbook No. 379, 1970.
56. G. C. Marten, G. E. Brink, D. R. Buxton, J. L. Halgerson, and J. S. Hornstein, *Crop. Sci.*, **24**: 1179 (1984).
57. K. A. Albrecht, G. C. Marten, J. L. Halgerson, and W. F. Wedin, *Crop. Sci.*, **27**: 586 (1987).
58. D. H. Clark, H. F. Mayland, and R. C. Lamb, *Agron. J.*, **79**: 485 (1987).
59. D. Clark, E. E. Cary, and H. F. Mayland, *Agron. J.*, **81**: 91–95 (1989).
60. R. G. Ward, G. S. Smith, J. D. Wallace, N. S. Urguhart, and J. S. Shenk, *J. Anim. Sci.*, **54**: 399–402 (1982).
61. D. D. Eckman, J. S. Shenk, P. J. Wangsness, and M. O. Westerhaus, *J. Dairy Sci.*, **66**: 1983–1987 (1983).
62. J. L. Holecheck, J. S. Shenk, M. Vavra, and D. Arthur, *J. Anim. Sci.*, **55**: 971–975 (1982).
63. J. S. Shenk, I. Landa, M. R. Hoover, and M. O. Westerhaus, *Crop Sci.*, **21**: 355–358 (1981).
64. S. M. Abrams, J. S. Shenk, M. O. Westerhaus, and F. E. Barton II, *J. Dairy Sci.*, **70**: 806–813 (1987).
65. F. E. Barton II, NIRS Workshop, ECOP Invit. Workshop on NIRS, Madison, WI, Nov. 13–15, 1985.
66. H. Mark and J. Workman, Jr., *Anal. Chem.*, **58**: 1454 (1986).
67. T. H. Blosser and J. B. Reeves III, *J. Dairy Sci.*, **69**: Abstr. Suppl., 136 (1986).
68. T. H. Blosser, High-Moisture Feedstuffs, Including Silage, in G. C. Marten, F. E. Barton II, and J. S. Shenk (Eds.), Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality. *USDA Agr. Hdbk.* No. 643, 1985, pp. 56–57.
69. K. H. Norris, USDA Research Center, Beltsville, MD, personal communication, August 12, 1987.
70. *Feedstuffs*, March 16, 1981, p. 43.
71. J. S. Shenk and M. O. Westerhaus, Population definition, sample selection, and calibration procedure for near infrared reflectance spectroscopy, *Crop. Sci.*, **31**: 469–474 (1990).
72. J. S. Shenk and M. O. Westerhaus, Population structuring of near-infrared spectra and modified PLS regression, *Crop. Sci.*, **31**: 1548–1555 (1990).
73. W. C. Templeton, Jr., J. S. Shenk, K. H. Norris et al., Forage Analysis with Near-Infrared Reflectance, Spectroscopy: Status and Outline of National Research Project, in *Proc. XIV Int. Grassl. Cong.*, G. J. Allen Smith and Virginia W. Hays (Eds.), Westview Press, Boulder, CO, 1983.
74. J. Workman, Near Infrared Spectroscopy: A Comparison of Random Versus Spectrally Selected Calibration Sets for Use in Forage Quality Evaluation, in *Proc. Forage and Grasslands Conf.*, Athens, Georgia, Am. For. Grass. Coun., Lexington, KY, pp. 132–136 (1986).
75. J. Shenk, Equation Selection, in *Near Infrared Reflectance Spectroscopy (NIRS); Analysis of Forage Quality*, G. C. Marten, F. E. Barton II, and J. S. Shenk (Eds.), *USDA Agr. Hdbk.* No. 643, pp. 26–27.

76. D. E. Honigs, G. M. Hieftje, H. L. Mark, and T. B. Hirschfeld, *Anal. Chem.*, **57**: 2299 (1985).
77. W. F. McClure, H. Aboul, F. G. Glaesbrecht, and W. W. Weeks, *Appl. Spectrosc.*, **38**: 222 (1984).
78. D. Bertrand, P. Robert, and V. Tran, Mathematical treatments of NIR spectra of mixtures, in *Proc. 11th I.C.C. Mtg.*, Vienna, 1985.
79. W. R. Hruschka and K. H. Norris, *Appl. Spectrosc.*, **36**: 261–265 (1982).
80. R. A. Isaac, Field Monitoring of Nitrogen in Plant Tissue by NIRA, in *Proc. Int. Symp. on NIRA*, Technicon Instr., Tarrytown, NY, 1982.
81. R. A. Isaac, Determination of protein in vegetative alfalfa and small grains using NIRA, in *Proc. Int. Symp. on NIRA*, Technicon Instr., Tarrytown, NY, 1983.
82. R. A. Isaac and W. Dorsheimer, *Am. Lab.*, April 1982.
83. P. Randall, Use of NIRA in animal feed analysis, in *Proc. Intl Symp. on NIRA*, Technicon Instrs., Tarrytown, NY, 1984.
84. E. V. Valdes, L. G. Young, I. McMillan, and J. E. Winch, Analysis of Hay, Haylage, and Corn Silage Samples by Near Infrared Reflectance Spectroscopy, in *Proc. Intl Symp. on NIRA*, Technicon Instrs., Tarrytown, NY, 1984.
85. E. V. Valdes, L. G. Young, I. McMillan, and J. E. Winch, Prediction of Chemical Composition of Hay, Haylage, and Corn Silage Samples by NIRA, in *Proc. Intl Symp. on NIRA*, Technicon Instrs., Tarrytown, NY, 1984.
86. E. V. Valdes, L. G. Young, and J. E. Winch, Application of near infrared reflectance spectroscopy (NIR) for the Evaluation of Forage Samples, in *Proc. Int. Symp. on NIRA*, Technicon Instrs., Tarrytown, NY, 1984.
87. W. C. Vercauteren, NIRA in compound feed production, in *Proc. Int. Symp. NIRA*, Technicon Instrs., Tarrytown, NY, 1984.
88. J. E. Winch, E. Valdes, and L. G. Young, The use of fixed wavelength N Infrared Analytical Equipment to Analyze Forages, in *Proc. Int. Symp. on NIRA*, Technicon Instrs., Tarrytown, NY, 1983.
89. J. Wolsink and H. Vedder, Experience with the InfraAlyzer 500 for Predicting Digestibility of Forages, in *Proc. Int. Symp. on NIRA*, Technicon Instrs., Tarrytown, NY, 1985.
90. Y. Park, M. Anderson, and J. Walters, *J. Dairy Sci.*, **66**: 235 (1983).
91. G. C. Marten, J. L. Halgerson, and J. H. Cherney, *Crop. Sci.*, **23**: 94 (1983).
92. G. C. Marten, I. Landa, M. R. Hoover, and M. O. Westerhaus, *Crop. Sci.*, **21**: 3 (1981).
93. J. H. Wolsink and H. M. Vedder, *Anim. Feed Sci. Technol.*, **15**: 190 (1986).
94. D. Burdick, F. E. Barton II, and B. Nelson, *Agron. J.*, **73**: 399 (1981).
95. J. M. deRuiter and J. C. Burns, *Crop. Sci.*, **27**: 1069 (1987).
96. P. C. Williams, *Cereal Chem.*, **52**: 561 (1975).
97. R. D. Benson, Statistical process control techniques using near-infrared reflectance spectroscopy, in *Proc. World Conf. on Emerging Technologies in the F and Oils Industry*, A. R. Baldwin (Ed.), American Oil Chemists Society, 1986.
98. D. J. Minson, K. L. Butler, N. Grummitt, and D. P. Law, *Anim. Feed Sci. Tech.*, **9**: 221–237 (1983).
99. D. E. Honigs, G. M. Hieftje, and T. B. Hirschfeld, *Appl. Spectrosc.*, **38**: 844 (1984).
100. H. Mark and J. Workman, Jr., *Spectroscopy*, **3**: 28–36 (1988).
101. J. S. Shenk and M. O. Westerhaus, Standardizing NIRS Instruments, in *Proc. Third Intl NIRS Conference*, Brussels, 1990.
102. M. O. Westerhaus, Improving Repeatability of NIR Calibrations across Instruments, in *Proc. Third Intl NIRS Conference*, Brussels, 1990.
103. J. Workman, Sample Selection Techniques, in *Proc. of the Third Intel NIRS Conference*, Brussels, 1990.
104. T. K. Nadler, S. T. McDaniel, M. O. Westerhaus, and J. S. Shenk, *Appl. Spectrosc.*, **43**: 1354–1358 (1989).
105. T. K. Nadler, Identifying Resonances in the Near-Infrared Absorption Spectrum, masters thesis, Pennsylvania State University, University Park, 1988.
106. H. Mark and J. Workman, Statistics in Spectroscopy, 2nd ed., Elsevier Science, Amsterdam, 2003.
107. J. S. Shenk, M. O. Westerhaus, and P. Berzaghi, Investigation of a LOCAL calibration procedure for near infrared instruments, *J. Near Infrared Spectrosc.*, **5**: 223–232 (1997).

108. P. Berzaghi, J. S. Shenk, and M. O. Westerhaus, LOCAL prediction with near infrared multi-product databases, *J. Near Infrared Spectrosc.*, 8: 1–9 (2000).
109. B. de la Roza, A. Martinez, S. Modrono, and B. Santos, Determination of the quality of fresh silage by near infrared reflectance spectroscopy, in *Near Infrared Spectroscopy: The Future Waves*, A. M. C. Davies and P. Williams (Eds.), pp. 537–541 (1995).
110. P. Dardenne, R. Agneessens, and G. Sinnaeve, Fresh forage analysis by near infrared spectroscopy, in *Near Infrared Spectroscopy: The Future Waves*, A. M. C. Davies and P. Williams (Eds.), pp. 531–536 (1995).

18 The Role of Near-Infrared Spectroscopy in Verifying Label Information in Agro-Forestry Products

Ana Garrido-Varo and Emiliano de Pedro

CONTENTS

18.1 Introduction	387
18.2 The Role of NIRS for the Labeling of Spanish Food Products with Special Characteristics	387
18.3 NIRS for Labeling Issues in the Animal Feed Industry	390
18.4 Traceability, Labeling, and Certification of Agro-Forestry Systems Products with NIRS and Other Sensors: The Dehesa Agro-Forestry System and the Traceability of the IP Products	392
18.5 Conclusions and Future	395
Acknowledgments	396
References	396

18.1 INTRODUCTION

There are numerous examples in the literature concerning the role of near-infrared spectroscopy (NIRS) for traceability, authentication, and labeling of agro-food products. There are also excellent review papers. Therefore, this chapter does not focus on an exhaustive literature review of the topic, but rather presents some examples, taken from the research experience of the authors, on the use of NIRS in mandatory and voluntary food and feed traceability and labeling programs.

18.2 THE ROLE OF NIRS FOR THE LABELING OF SPANISH FOOD PRODUCTS WITH SPECIAL CHARACTERISTICS

In Spain there are many food products (wine, spirits, cheeses, virgin olive oils, cured ham, rice, dried pulses, peppers, asparagus, fruits, picked vegetables, fresh meat, cured sausages, honey, nougat, and cured meats) that are protected under a Protected Designation of Origin (PDO) or a

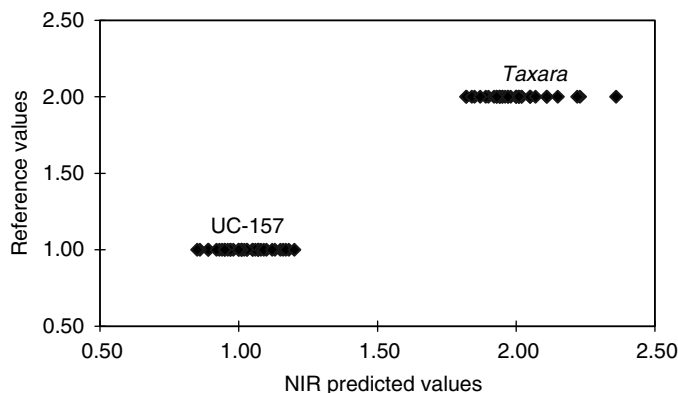


FIGURE 18.1 Reference values vs. NIR predicted values corresponding to 74 asparagus samples used for independent validation of the prediction model.

Protected Geographical Indication (PGI). PDOs and PGIs are systems of quality labels. A PDO can be attributed to foodstuffs that are produced, processed, and prepared in a given geographical area using recognized know-how. In the case of the PGI, the geographical link to a specific place or region must be in at least one of the stages of production, processing, or preparation. The characteristics and authenticity of these protected products are controlled by the respective Regulating Councils.

One example of a PGI may be found in the green asparagus “Huetor Tajar,” which is produced in a small area of Granada (Southern Spain). The corresponding Regulating Council warrants that the asparagus is from one autochthonous variety (i.e., Taxara). We have demonstrated that a partial least squares (PLS) regression (Figure 18.1) using as variables 1 = UC-157 (hybrid variety) and 2 = Taxara (autochthonous variety) allows a clear differentiation between the two asparagus classes [1].

Iberian ham is probably the most expensive luxury food product produced in Europe. It is obtained from pigs left free to graze on grass and acorns in Spanish *dehesas* (agrosilvopastoral systems). The type of feeding during the final growing period produces hams classified in three commercial categories: “acorn” (125 €/kg), “recebo” (60 €/kg), and “concentrates” (25 €/kg). The quality and uniqueness of the Iberian pig (IP) hams are guaranteed by the appropriate PDOs Regulating Councils by on-farm inspection and gas chromatography (GC) analysis of the fatty acids (FA) of the adipose tissue (C16:0, C18:0, C18:1, and C18:2). The cost and time to obtain FA analyses by GC means that, in practice, only a mean sample of fat coming from many animals is analyzed by GC, and these values are used to classify the entire lot. Traceability and labeling of each individual ham is of great value both for industrials and consumers. At the NIR Conference held in Norway in 1991, our research group presented the preliminary results of the evaluation of NIRS technology for predicting the four main FAs present in IP fat. We later demonstrated that equations produced by using transmission, transreflectance, and interactance–reflectance have a similar accuracy and precision. Furthermore, calibration equations using transreflectance has been successfully transferred between instruments [2,3].

At present, NIRS has been implemented in laboratories and slaughterhouses for the classification of each individual carcass using different analysis modes. Figure 18.2 shows the methodology implemented, at industry level, to pay farmers according to the three Iberian pork commercial categories (“acorn,” “recebo,” and “concentrates”). The methodology consists of sampling the adipose tissue of a carcass, which is then analyzed at room temperature using an NIR instrument with an attached interactance–reflectance fiber optic probe. Once the sample spectrum is obtained, NIR calibration equations are used to predict instantaneously the fatty acid content of the sample [4].

Olive oil is the only commodity in the fat and oil sectors to have its own international agreement. The European directives governing the characterization of olive oil define 24 analytical characteristics

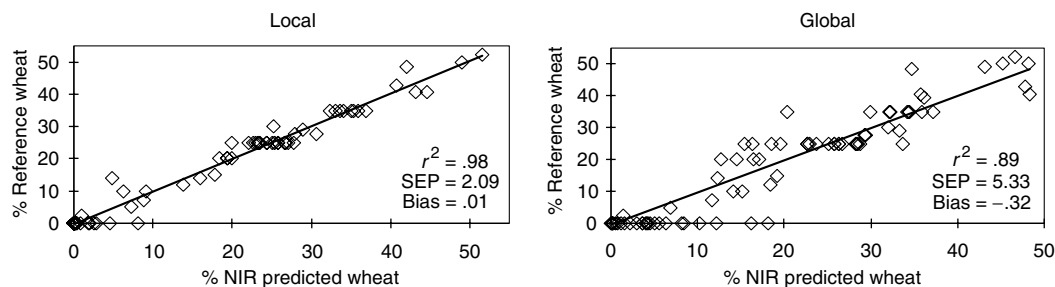


FIGURE 18.2 Prediction of percentage of wheat in validation set samples using GLOBAL vs. LOCAL equations.

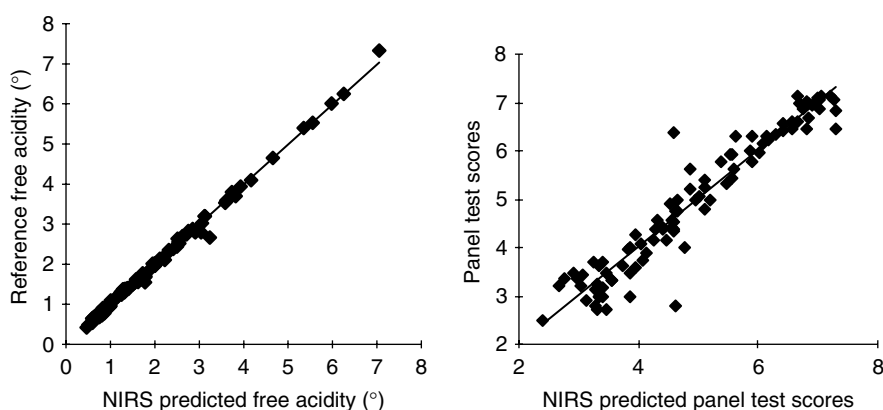


FIGURE 18.3 Predicted vs. reference values for $n = 105$ samples of olive oils for free acidity and panel test scores.

of these oils, enabling olive oil to be classified into nine different categories:

1. Extra virgin olive oil
2. Virgin olive oil
3. Ordinary virgin olive oil
4. Lampante virgin olive oil
5. Refined olive oil
6. Olive oil
7. Crude olive-pomace oil
8. Refined olive-pomace oil
9. Olive-pomace oil.

In practice, for commercial transactions between producers, superstores, supermarkets, delicatessens shops, or restaurants, it is standard practice to provide at least information on free acidity or free fatty acid (FFA), peroxide value (PV), specific extinctions at 232 nm (K_{232}) and 270 nm (K_{270}), and panel test scores (PTS).

During the NIR-99 Conference held in Verona, we presented results of NIR calibration equations for the prediction of FFA, PV, specific extinctions at 270, 225, and 232 nm (K_{270} , K_{225} , and K_{232}), oxidative stability (S), moisture (H), polyphenols (Pol), and PTS. The results showed that NIR spectroscopy could explain the 99, 72, 77, 88, 65, 92, 76, 93, and 86% of the variation existing in FA, PV, K_{270} , K_{225} , K_{232} , S, H, Pol, and, PTS respectively. Figure 18.3 shows the predicted vs.

reference values for $n = 105$ samples of olive oils for FA and PTS. It is clear that extra virgin olive oil ($FA < 1$ and $PTS \geq 6.5$) can be clearly distinguished from other categories using NIRS.

It is well known that if we want to have high-quality virgin olive oil, we need to produce *olive fruits* of high quality. Therefore, olive-oil processing plants are increasing the quality control of the olive fruits used as raw material to produce virgin olive-oil. Free acidity and fat yield are the most relevant chemical parameters for a quick and reliable quality control and payment of olive fruits at the intake of the olive-oil mill plant. We have demonstrated that intact olive fruits can be analyzed by NIR and the free acidity and fat content can be predicted, with high precision and accuracy. The equations explained 91 and 96% of the variations in free acidity and fat content, respectively. The relative percent difference (RPD) values or the ratio of the error of performance for cross-validation (SECV, standard error of cross-validation) to the standard deviation (SD) of the reference data were 3.3, and 4.6, which indicate the possibility of accurate analysis by using the equations obtained [5].

18.3 NIRS FOR LABELING ISSUES IN THE ANIMAL FEED INDUSTRY

During the last few years we have concentrated great efforts on demonstrating the role of NIR for labeling issues related to the detection and quantification of ingredients in compound feeds (CFs) and animal products (protein meals and fats). Light microscopy, currently the official method for identifying ingredients, has certain well-documented limitations when applied to quantitative analysis; moreover, it is a time-consuming technique and therefore unsuited to the routine quality-control tests required for the animal feed industries under the new EU legislation.

Concerning CFs, our research was aimed at demonstrating that the NIR analysis of ground and intact CFs produces accurate information not only on traditional and compulsory labeling constituents (moisture, protein, fat, ashes, and fiber), but also on other unconventional constituents (Table 18.1) such as the ingredient composition, especially meat and bone meal (MBM), and other ingredients that became items for compulsory declaration in November 2003 [6]. We have demonstrated that the equations to predict the ingredient percentage had similar accuracy for the analysis of ground and intact compound feeding stuffs [7]. The equations showed an excellent ability ($r^2 \geq 0.9$; $RPD \geq 3$) to predict the inclusion percentage of ingredients such as sunflower meal, gluten meal, lucerne, beet

TABLE 18.1
Calibration Statistics for Predicting the
MBM Percentage of Ground and Intact
Compound Feeds

	% MBM	
	Ground set	Intact set
<i>N</i>	1005	523
Mean	3.24	3.09
Range	0.00 to 34.85	0.00 to 32.55
SD	5.50	5.23
SECV	0.94	0.80
r^2 CV	0.97	0.98

N = number of samples; SD = standard deviation;
SECV = standard error of cross-validation; r^2 =
coefficient of determination for cross-validation.

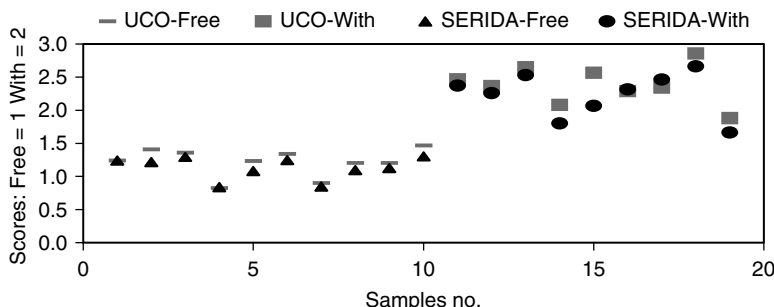


FIGURE 18.4 External validation ($n = 18$) of a discriminant model for the prediction of the presence/absence of MBM. NIR analyses in two cloned instruments located in two different laboratories (UCO and SERIDA).

pulp, palm meal, MBM, poultry meal, total meat meal (“MBM + poultry meal”), animal fat, and mineral-vitamin supplement. However, for other ingredients (e.g., cereals) the standard error of prediction (SEP) values were high.

Latter on, in close cooperation with the NUTRECO group, it was demonstrated that by using large library files (7423 CF samples) and the nonlinear chemometric algorithm LOCAL (WINISI software ver. 1.05, InfraSoft International, LLC., Port Mathilda, USA), the equations for all the ingredients greatly improved in relation to those obtained with Global PLS equations. For example, Figure 18.2 shows the reduction in the SEP value obtained (5.33 to 2.09%) for the prediction of wheat percentage when using the LOCAL algorithm [8,9].

The ban on the use of animal meals in CFs [10] is one of the measures that have been taken in the EU to stop the spread of bovine spongiform encephalopathy (BSE) and to prevent its re-emergence.

With respect to the ingredient MBM, we have developed both quantitative and qualitative NIRS models, for the prediction of the percentage of MBM and also for the detection of the presence or absence of MBM in CFs [6].

The results obtained until now demonstrate that qualitative discriminant model produces fewer wrong classified samples as free or containing MBM. A qualitative/discriminant model ($N = 523$ intact CFs, $r^2 = .86$, $SECV = 0.19$) has been validated with 18 CF samples representative of those circulating at present through Europe (ten CFs for production animals free of MBM and eight for pet foods containing MBM). They were analyzed in two cloned instruments. As can be seen in Figure 18.4, CFs free of MBM group together in one cluster around a score of 1.0 for the same samples scanned in the two instruments, while CFs containing different levels of MBM cluster around 2.0. With a breakpoint score of 1.5, all the samples were discriminated correctly by the two instruments.

Inexpensive, fast, reliable, and automated methods are needed for the enforcement of legislation concerning animal by-products (protein, animal meals, and fats), such as the EU Regulation EC No. 1774/2002 [10] governing Animal Processed By-Products (ABPs), which addressed the possible risk inherent in recycling potential BSE infectivity due to the absence of barrier within species and claimed to exclude the cannibalism, which may be induced by the intraspecies recycling. The Regulation opened discussions about the need for analytical methods to differentiate among the different categories of animal meals and fats.

In close cooperation with rendering plants, we are developing spectral libraries and databases with useful traceability information such as the percentage of meat of each animal species (i.e., pig, cow, poultry, ewe, etc.) used to produce the rendered meal or fat.

Figure 18.5 shows the mean spectrum of pure poultry meals ($n = 56$), pure pork meals ($n = 16$), pure cattle meals ($n = 1$), cattle-poultry meal mixture ($n = 1$), cattle-pork meal mixture ($n = 3$), poultry-pork meal mixture ($n = 3$), blood meal ($n = 1$), fish meal ($n = 8$), hydrolyzed feather meal ($n = 1$), feather meal ($n = 1$), and cattle-sheep-pork and poultry meal mixtures ($n = 189$). There are visible differences in the region 1680 to 1760 nm that may be the basis for differentiating among the different types of protein animal meals [11].

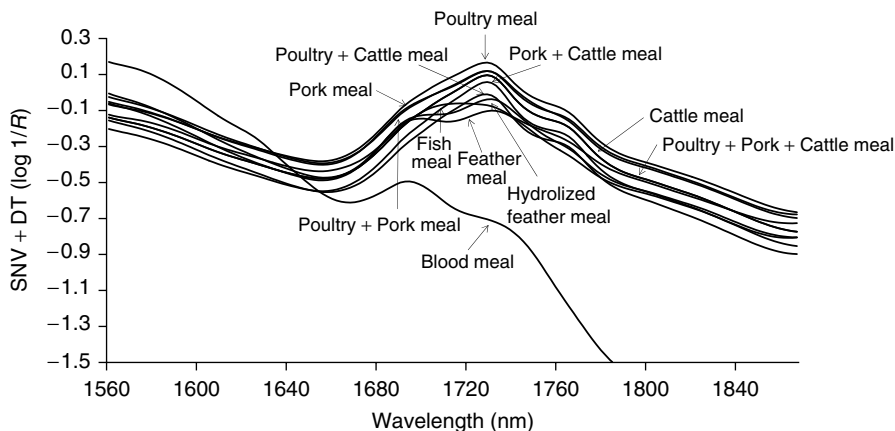


FIGURE 18.5 Relevant absorption features of the mean spectrum of different animal protein meals.

The results obtained so far show the potential of NIR technology to identify the animal species in animal by-products. Further work is in progress to enlarge our spectral library files and to evaluate different quantitative and qualitative chemometric strategies to build models that can be applied to all the ABPs (animal meals and fats) circulating at intra- and inter-European Community level.

18.4 TRACEABILITY, LABELING, AND CERTIFICATION OF AGRO-FORESTRY SYSTEMS PRODUCTS WITH NIRS AND OTHER SENSORS: THE DEHESA AGRO-FORESTRY SYSTEM AND THE TRACEABILITY OF THE IP PRODUCTS

Agro-forestry system is a term used to describe land management systems that combine the production of trees, with crops and/or animals preserving local practices to produce high value-added food and nonfood products. The “Dehesa” is one example of a complex agro-forestry system also called silvopastoral, and it is the natural habitat of the IP. The Dehesa is considered a multiple resource production system based on the harvesting of a variety of products such as meat, milk, eggs, cereals, honey, olive fruits, cork, charcoal, timber, defence against fire, water, landscape, diversity, integration of domestic and wild fauna, shelter for fauna, and so forth. A well-managed Dehesa is one of the best examples of integration between grassland productivity and sustainability [12,13].

The optimum quality of the IP hams (“acorn” category) is reached by keeping a given production system that is mainly based in the breed purity of the animals (at least 75% of IP blood) and consumption of Dehesa natural resources. In other words, during the final growing period (165 to 175 kg; each individual pig must have gained 575 kg) or at least 60% of the weight at the start of the finishing period (86 to 105 kg) by consumption of acorns (fruits of the *Quercus ilex* spp), grass and other natural resources [14]. However, variable agro-climatic conditions cause enormous year-to-year variability in the available natural resources and thus in the number of animals belonging to each of the three commercial categories: “acorn,” “recebo,” and “concentrates” (animals only feed with CF).

IP hams have a high national and international reputation not only because of its taste and flavor, but also because of health properties (e.g., high oleic acid content) and because of the “organic” nature of the IP production system. At present, the inspection controls established for the IP are based upon “on farm” inspector visits and fatty acid analysis done in one fat sample taken from a group of animals from the same producer. That system is very time consuming and expensive, and not objective enough to fulfil the increasing consumer demands from regional, national, and

international markets. Another downside of the system is that it does not provide information about the production and quality of each individual ham.

The implementation of traceability, labeling, and certification systems is rather more complicated in agro-forestry systems than in industrial production processes. In fact, the understanding and definition of the scientific and technological basis for such control systems override individual capabilities of companies and suppliers that offer traceability systems. However, for the Dehesa, and in particular for the IP production system, it is crucial to develop such systems to avoid frauds and to ensure consumer expectations after paying for such a luxurious and expensive product because not only does it taste delicious but also because of its contribution to the consumer's health and to the environment.

At present, new methodologies, technologies, and research trends are emerging that provide a multidisciplinary approach for building more innovative, creative, sustainable, credible, and affordable traceability and labeling programmes. Figure 18.6 represents a proposal for the use of different spectral complementary sensors placed in different locations and supports (laboratories, hand-held, aeroplanes, helicopters, satellites) for the management of the Dehesa. Although that figure represents a system applicable to all the Dehesa products (grass, honey, milk, cork), we will focus our comments on the management of the IP production system.

The set of tools and utilities shown in Figure 18.6 will allow one to record critical and essential data for the sustainable production and traceability of the ham and other expensive IP products (e.g., shoulders, loins, and cold meats) and it will represent a significant step forward in our current strategy of evaluating the quality based on NIRS (Figure 18.7) because the control will begin in the Dehesa (Figure 18.6 and Figure 18.8).

The system shown in Figure 18.6 and Figure 18.8 will have the following components:

1. Sampling of available feeding (mainly grass and acorns) on selected Dehesa sites using georeferenced positioning systems.
2. Laboratory analysis of the samples by wet chemistry, genomics, and NIRS. These analyses will provide data of the many components of the Dehesa (grass, acorns, soil, leaves). Then,

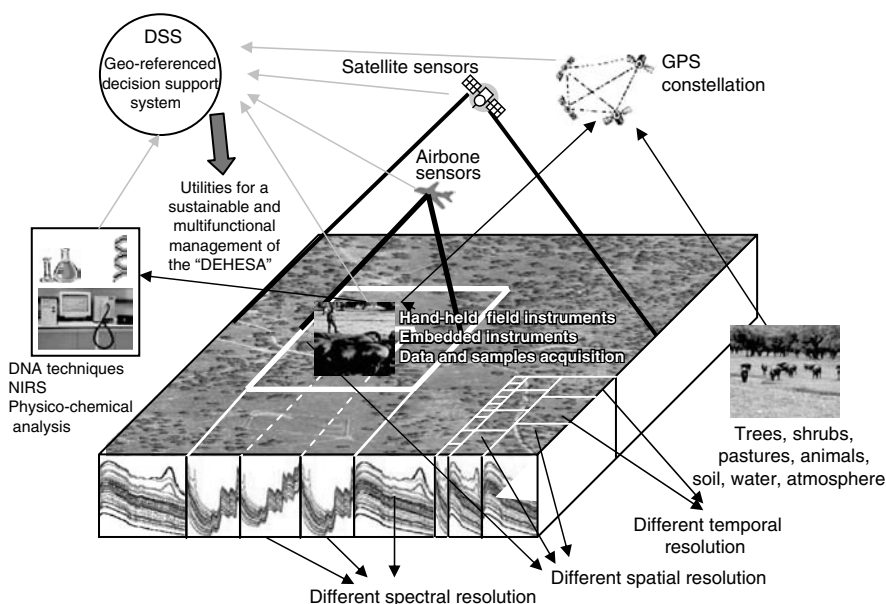


FIGURE 18.6 Tools and utilities for a sustainable and multifunctional management of the “Dehesa.”

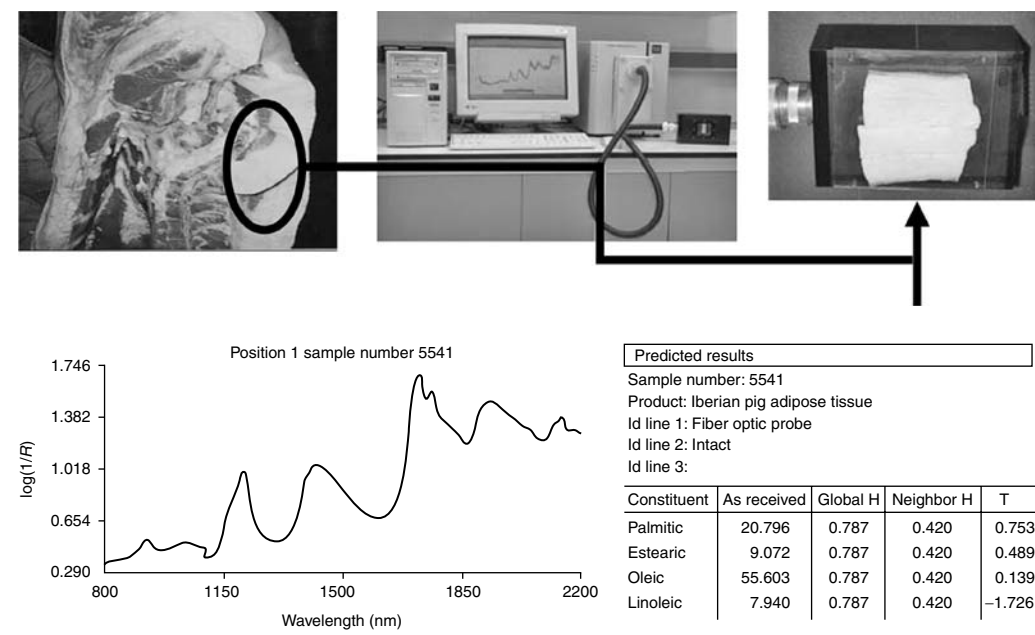


FIGURE 18.7 Sampling and NIRS analysis by interactance-reflectance. According to the four fatty acids each carcass is classified in one of the three commercial categories and the farmer is paid accordingly.

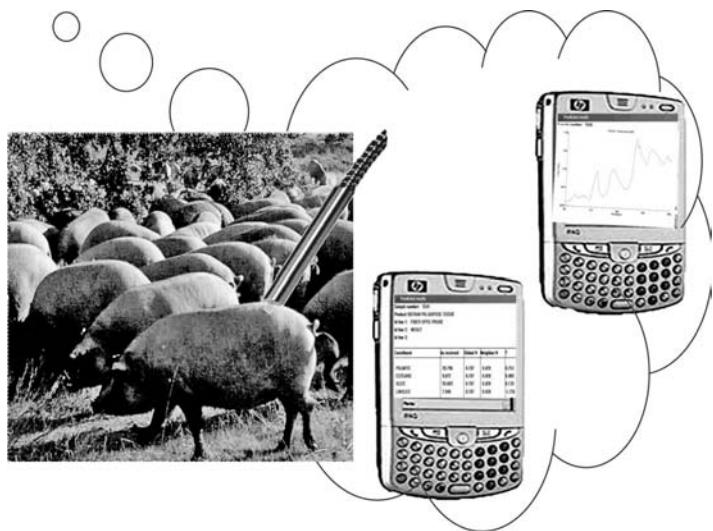


FIGURE 18.8 Future on-farm management of Iberian pig feeding based on NIRS.

- these data will be used as inputs for hand-held sensors (e.g., NIRS, radiometers, etc.) and aero-transported and satellite sensors.
3. Calibration of on-farm VIS/NIRS instruments to predict quality and quality of the available feed (e.g., botanic composition of the grass, digestibility, FA of the acorns, fruits, etc.).
 4. Hyperspectral sensors mounted in small airplanes, helicopters, and satellites.
 5. Calibration of aero-transported VIS/NIRS and hyperspectral satellite instruments with data obtained at field levels.

6. Modeling methods to produce qualitative and quantitative new approaches to predict critical information (e.g., amount and quality of the available grass and acorns, expected live weight gain, etc.). The models should be achievable by exploiting the complementary features of the information produced by different sensors.
7. Dynamic mapping of the seasonal and annual available natural feeding resources existing in the Dehesa.
8. A Global Positioning System (GPS) attached to each individual pig.
9. VIS/NIRS spectral data obtained from each individual pig, using a portable instrument provided with a fiber optic probe. Instantaneous prediction of the fatty acid profile (Figure 18.7).
10. Input of the fatty acid profile and other data of interest (GPS coordinates, weight, breed purity, DNA traceability using PCR, etc.) to the Decision Support System (DSS).
11. The DSS should not only be designed for supporting management decisions but also for use as a tool to connect, store, and organize a network knowledge base built from results of isolated and (hopefully) integrated and multidisciplinary Dehesa projects.

In Spain, the management system proposed in Figure 18.6 is currently under development in the framework of different R&D and innovation projects.

Thus, the University of Extremadura together with private communication companies is running a joint project to observe pigs remotely. Each individual pig is fitted with a transmitter. The transmitter emits a GPS signal that is sent to a General Packet Radio Service (GPRS) network [15,16].

Researchers from the National Institute for Agronomical Research (INIA) have conducted projects in the field of the DNA mapping of the IP breed [17].

Researchers from the University of Córdoba are working on several projects aimed at reducing the increasing degradation of the Dehesa due to different factors such as inadequate pruning, overgrazing, acorn tree pathologies, and so forth [18]. The projects are also oriented to study different strategies and methodologies for Dehesa inventory and regeneration [19]. The tools and utilities developed in these projects, such as a Geographical Information System (GIS) containing data such as type of soil, tree variety, acorn weight, acorn quality, and so forth, also need to be integrated into the Dehesa management system represented in Figure 18.7.

Another group at the University of Cordoba, in close cooperation with different departments from the Government of Andalusia (Agriculture, Environment, Agro-Food Research and Training Institute), is working on the development of methodologies, supported in GIS, remote sensing, photogrammetry, and airborne laser technology, to obtain semiautomatic information of the architecture of agro-forestry systems (e.g., number of trees and tree topology).

Years of research conducted by this chapter's authors have demonstrated that the IP pig subcutaneous outer back fat, which is the layer immediately beneath the skin, contains a different proportion of the four major FAs (C16:0, C18:0, C18:1, and C18:2) depending on the feeding regime. They also demonstrated that the FA profile of each individual animal can be accurately predicted by NIRS, either using melted fat or intact adipose tissue [20–27]. A project has recently been started aimed at evaluating different portable instruments to be used in the field as shown in Figure 18.8. Calibration equations developed with that type of instrument will provide useful information about the animal and the available feed that then will be used as input to other complementary sensors and to the DSS.

18.5 CONCLUSIONS AND FUTURE

NIRS has demonstrated its ability to be integrated in traceability, quality, and labeling programs in the agro-food sector. However, NIRS is not the only sensor needed for management and traceability of complex Agriculture and Forestry farming systems.

The strategy described here for Dehesa management with the integration of different complementary sensors will offer the possibility of collecting information data in an agile, dynamic, and less costly manner and with different spatial, temporal, and spectral resolutions.

The increasing promotion of regional, national, and international multidisciplinary projects in agro-forestry management will facilitate the integration of research activities that, at present, are being undertaken isolated, and will contribute to reaching the proposed strategy for a sustainable and multifunctional management of the “Dehesa” and of the IP products.

ACKNOWLEDGMENTS

The work described in this chapter would not have been possible without the collaboration of many of our pregraduate and postgraduate students, laboratory technicians, and representatives of several agro-food industries. We are especially grateful for the contributions of the following persons: Prof. Dr. José E. Guerrero Ginel for his ideas and suggestions for completing the manuscript, Prof. Dr. Augusto Gómez-Cabrera, Dr. M.D Pérez-Marin, Dr. Nieves Nuñez. Dr. Juan Garcia-Olmo, MSAE Maria José de la Haba, MSAE Leo Ortiz, Ms. Francisca Baena, Mr. Antonio López, Mr. Manuel Sánchez, and Ms. María Luisa Mármol for contributions to the research reported in this chapter. Thanks are also due to the following Spanish industries and laboratories: SAPROGAL, RENDER-SUR, CORSEVILLA, NUTRECO, and Laboratorio Agroalimentario de Córdoba. Finally, A. Garrido thanks Dr. Burns for his invitation to contribute to this book and for his suggestions for preparing the manuscript.

REFERENCES

1. Pérez, M.D., Sánchez, M.T., Cano, G., and Ana Garrido, A. Authentication of green asparagus varieties by near-infrared spectroscopy. *Food Science*, 6: 323–327 (2001).
2. García-Olmo, J., Garrido, A. and De Pedro, E. The transfer of fatty acids NIRS calibration equations using four sets of unsealed liquid standization samples. *J. Near Infrared Spectrosc.*, 9: 49–62 (2001).
3. Garrido, A., Garcia-Olmo, J. and Pérez-Marin, M.D. Application in the analysis of fat and oils. In: *NIR Spectroscopy in Agriculture*. Cap. XIX. C. Roberts, J. Workman and J. Reeves (Eds.). American Society of Agronomy (ASA), Crop Science Society of America (CSSA) and Soil Science Society of America (SSSA), pp. 487–558 (2003).
4. De Pedro, E., Nuñez, N., García, J. et al. Implementing Near Infrared Spectroscopy in The Iberian Pork Industry. *5th International Symposium of the Mediterranean Pig*. Tarbes, France (2005).
5. González, J.M., Pérez-Marín, M.D., Garrido-Varo, A. and Guerrero, J.E. Non-destructive analysis of intact olive fruits by near infrared spectroscopy prediction of fat yield and acidity. In: *NIR 2003*. T. Davies and A. Garrido (Eds.). NIR Publications, Chichester, UK, pp. 373–376 (2004).
6. European Commission (EC). Directive 2002/2/EC of the European Parliament and of the Council. OJ L 63, 23 (2002).
7. Pérez-Marín, D., Garrido-Varo, A., Guerrero-Ginel, J.E. and Gómez-Cabrera, A. Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: chemical composition and open-declaration. *Animal Feed Science and Technology*, 116: 333–349 (2004).
8. Pérez-Marín, D., Garrido-Varo, A., Guerrero, J.E. et al. Ad Hoc labelling of compound feedingstuffs: the feed industry bets by near infrared spectroscopy. *NIR News*, 14: 3–5 (2003).
9. Pérez-Marín, D., Garrido-Varo, A., Guerrero-Ginel, J.E. and Gómez-Cabrera, A. Implementation of local algorithm with NIRS for compliance assurance in compound feedingstuffs. *Applied Spectroscopy*, 59: 69–77 (2005).
10. European Commission. Decision 2000/766/EC of December 4. OJ L 306, 32–33 (2000).
11. European Commission. Regulation EC N 1774/2002 OJ L 273, 1–95 (2002).
12. Garrido-Varo, A., Pérez-Marín, M.D., Guerrero, J.E. et al. Near infrared spectroscopy for enforcement of European Legislation concerning the use of animal by-products in animal feeds. *Biotechnol. Agron. Soc. Environ.*, 9: 3–9 (2005).

13. Gastó, J., Remmers, G.G.A., Fernández, P., Guerrero, J.E and Sevilla, E. Agricultural landscape production at the farming level in the Mediterranean environment of Spain. Paper presented at the first plenary meeting of the EC Concerted Action "The landscape and nature Production Capacity of Organic/Sustainable Types of Agriculture in the EC". Wageningen Agricultural University, Wageningen, November 29 to December 1, 1993, p. 18.
14. European Grassland Federation. Sustainable Grassland Productivity. Retrieved March 23, 2006 from the World Wide Web: <http://www.egf2006.com/location/location.htm> (2006).
15. B.O.E. Real decreto 1083/2001 de 5 de octubre por el que se aprueba la norma de calidad para el jamón ibérico, paleta ibérica y caña de lomo ibérico elaborados en España. BOE num. 247, lunes 15 de octubre (2001).
16. Aparicio, M.A., Vargas, J.D. and Atkinson, A. Las nuevas tecnologías y la montanera del cerdo ibérico. *Mundo Ganadero*, 185: 42–48 (2006).
17. Walmsley, K. Developer helps deliver high-tech hams. Retrieved March 23, 2006 from the World Wide Web: http://www1.ericsson.com/mobilityworld/sub/articles/other_articles/06jan3 (2006).
18. Ovílo, C., Barragán, M.C., Castellanos, C., Rodríguez, M.C., Silió, L. and Toro, M.A. Application of molecular markers (RAPD, AFLP and Microsatellites) to Iberian pig genotype characterization. In: *Tradition and innovation in Mediterranean pig production = Tradition et innovation dans la production porcine méditerranéenne*. J.A. Alfonso de Almeida and J.L. Tirapicos Nunes (Eds.). Zaragoza. CIHEAM-IAMZ, pp. 79–84 (2000).
19. Navarro Cerrillo, R. and Fernández Rebollo, P. El síndrome de la seca del encinar. Fundación Ricardo Delgado Vizcaino. Pozoblanco (Córdoba). Tipografía Católica SCA. Córdoba, Spain, p. 172 (2000).
20. Gómez-Cabrera, A. and De Pedro, E. La sostenibilidad de la Dehesa. Paper presented at the 21st General Meeting of the European Grassland Federation, Badajoz, Spain, April 3–6 (2006).
21. De Pedro, E., Garrido, A., Bares, I., Casillas, M. and Murria, I. Application of near infrared spectroscopy for quality control of Iberian pork industry. In: *Near Infrared Spectroscopy: Bridging the Gap between Data Analysis and NIR Applications*. K.I. Hildrum et al. (Eds.). Ellis Horwood, Chichester, UK, pp. 345–348 (1992).
22. Hervás, C., Garrido, A., Lucena, B., García, N. and De Pedro, E. Near infrared spectroscopy for classification of Iberian pig carcasses using an artificial neural network. *Journal of Near Infrared Spectroscopy*, 2: 177–184 (1994).
23. De Pedro, E., Garrido, A., Lobo, A., Dardenne, P. and Murria, I. Objective classification of Iberian pig carcasses: GC versus NIRS. In: *Leaping Ahead with Near Infrared Spectroscopy*. G.D. Batten et al. (Eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia, pp. 291–295 (1995).
24. García-Olmo, J., Garrido, A. and De Pedro, E. Advantages and disadvantages of MLR and PLS regression equations for the prediction of fatty acids. In: *Near Infrared Spectroscopy: Proceedings of the 9th Int. Conference*. A.M.C. Davies and R. Giangiacomo (Eds.). NIR Publications, Chichester, UK, pp. 253–258 (2000).
25. De Pedro, E., Nuñez, N., Garrido, A. et al. Qualitative analysis of NIRS spectral data to identify Iberian pig feeding types. In: *Eurofins Scientific (Ed.) Proceedings of the 6th Intl Symposium of Food Authenticity and Safety*. Eurofins Scientific, Nantes, France, p. 12 (2001).
26. De Pedro, E. Calidad de las canales y de los productos derivados del cerdo ibérico. Técnicas de control y criterios de calidad. In: *Porcino Ibérico: Aspectos Claves*. C. Buxadé (Ed.). Mundiprensa, Madrid, pp. 589–621 (2001).
27. García-Olmo, J., Garrido, A. and De Pedro, E. The transfer of fatty acids calibration equations using four sets of unsealed liquid standardisation samples. *Journal Near Infrared Spectroscopy*, 9: 49–62 (2001).
28. Garrido-Varo, A., García-Olmo, J. and Pérez-Marín, M.D. Applications in fats and oils (Chapter 19). In: *NIR Spectroscopy in Agriculture*. C. Roberts, R. Windham and J. Workman (Eds.). American Society of Agronomy (ASA), Crop Science Society of America (CSSA) and Soil Science Society of America (SSSA), pp. 487–558 (2004).

19 NIR Analysis of Cereal Products

B. G. Osborne

CONTENTS

19.1	Ingredients	399
19.1.1	Flour.....	399
19.1.1.1	Protein	400
19.1.1.2	Moisture	401
19.1.1.3	Particle Size	401
19.1.1.4	Ash	402
19.1.1.5	Color	402
19.1.1.6	Starch Damage and Water Absorption	403
19.1.2	Baking Quality of Flour	404
19.1.3	Wheat Gluten	405
19.1.4	Additive or Nutrient Premixes	405
19.1.5	Packaging Materials	406
19.2	Dough	406
19.3	Cereal Foods	407
19.3.1	Bread	408
19.3.2	Biscuits.....	409
19.3.3	Flour Confectionery	409
19.3.4	Breakfast Cereals	410
19.3.5	Pasta.....	410
19.3.6	Extrusion Cooking.....	410
19.3.7	Miscellaneous Cereal Foods	411
19.4	Conclusion	412
	References	412

This chapter reviews the applications of near-infrared (NIR) spectroscopy in the cereals industry for the analysis of ingredients (especially flour), process intermediates (dough), and final products (bread, biscuits, flour confectionery, pasta, and breakfast cereals).

19.1 INGREDIENTS

19.1.1 FLOUR

The most important ingredient of wheat-based products is flour, which is therefore the subject of quality specifications. The analysis of flour is the major practical in-bakery application of NIR. The experimental simplicity of NIR means that nontechnical bakery personnel easily operate the equipment. Its capability of generating essentially instantaneous results allows it to be used before

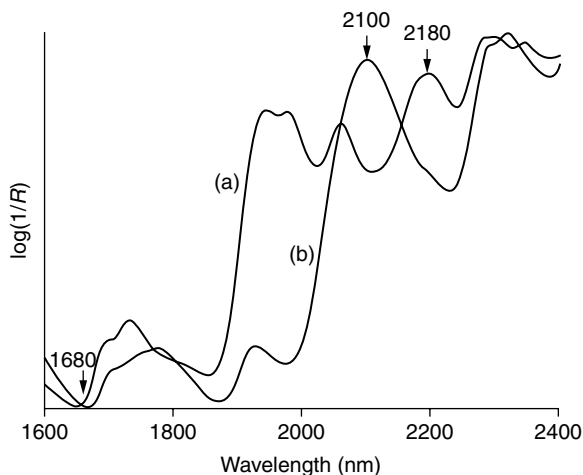


FIGURE 19.1 NIR spectra of wheat gluten (a) and wheat starch (b).

unloading bulk flour into the bakery's silos. Thus, flour shipments may be checked for conformity to specifications, so those unsatisfactory loads may be rejected before they enter the production system and compromise the quality of the bakery's product line. The flour quality parameters that have been determined by NIR are protein, moisture, particle size, ash, color, starch damage, and water absorption.

19.1.1.1 Protein

Protein is the most important single constituent in bakers' flour and has therefore received the most attention from NIR researchers [1–7]. Protein determination would normally be carried out on a filter instrument containing a limited number (between 6 and 19) of filters at fixed wavelengths. These data are fitted by least squares to protein values determined by the Kjeldahl or Dumas method. For example, Osborne et al. [3] employed a three-term equation involving $\log(1/R)$ at 2180, 2100, and 1680 nm. This can be seen to be a sensible equation because 2180 nm corresponds to an absorption band in wheat gluten that lies on the shoulder of a broad starch band with an absorption maximum at 2100 nm (Figure 19.1). Therefore, 2180 nm is the primary protein measurement wavelength, while 2100 nm allows a correction to be made for the effect of starch at 2180 nm. However, the 2100 nm term is involved in the calibration also because there is a high negative intercorrelation between starch and protein that is a characteristic of flour, so that a measurement at a wavelength characteristic of starch adds constructively to the calibration.

The accuracy of calibrations for the determination of protein in straight-grade flour when assessed on samples differing from those used in the calibration is of the order of 0.20% expressed as a standard error of prediction (SEP) [1,3,5]. Further, the calibrations have been found to be applicable to a wide range of sample types [1,3] and stable over a long period of time without adjustment [3]. Because NIR is being used as a practical tool to check bakery flour specifications [2], it is inevitable that disputes will arise between the supplier and the baker as to the accuracy of the NIR calibration. The usual course of action in such circumstances would be to analyze the disputed sample by the reference method, when almost certainly a discrepancy will be found between this result and the NIR result. This does not necessarily mean, however, that the calibration requires adjustment. The reason is to be found by partitioning the variance defined by the SEP [5].

The derivation and validation of an NIR calibration requires a great deal of time and effort. It is of enormous benefit for the same calibration to be used in other instruments of the same model, so that the entire procedure need not be repeated in each bakery. It is an important prerequisite,

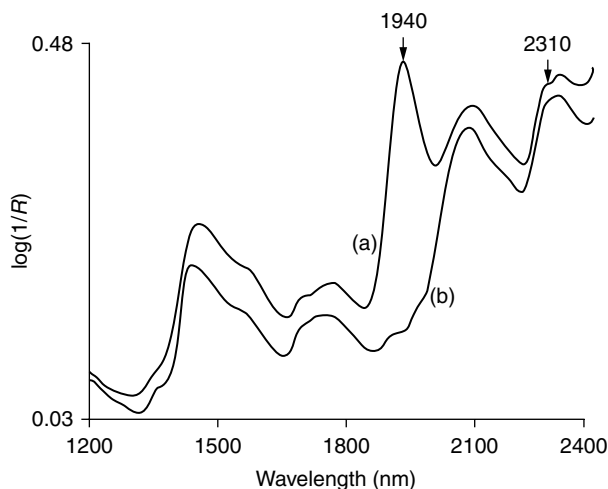


FIGURE 19.2 NIR spectra of a flour sample before (a) and after (b) oven-drying.

however, that the calibration be universal for all the flour types to be analyzed in all the bakeries in which it will be used. Transfer of calibration constants for protein in flour between nine fixed filter instruments has been demonstrated by Osborne and Fearn [6]. The instruments were situated in different laboratories and the calibration intercept for each was adjusted using a set of 20 samples. A second set of 20 samples was used to evaluate the performance of the instruments compared with the Kjeldahl method. The mean of the NIR results varied from 10.92 to 11.30%, while the mean Kjeldahl result was 11.03%. Analysis of variance gave a reproducibility of 0.17% for NIR and 0.14% for Kjeldahl. Regular monitoring of both NIR and Kjeldahl results may be carried out by means of check samples [8].

19.1.1.2 Moisture

One of the strengths of the NIR technique is its ability to provide the simultaneous determination of several constituents on the same sample. The determination of moisture was the first application of NIR in food analysis, and its importance warrants a whole chapter in this book. It is sufficient here to mention that the moisture content of bakers' flour needs to be determined and that this may be accomplished concurrently with protein [3,7]. Moisture has a strong absorption band at 1940 nm that is not overlapped by bands due to other constituents of flour (Figure 19.2); therefore calibration is straightforward because only a single reference wavelength (2310 nm) is required [9].

The accuracy of NIR calibrations for moisture SEP is of the order of 0.15% [1,3,5], which is better than for protein. Between-laboratory performance [6] is also better with mean values for 20 samples ranging from 13.08 to 13.28% between nine instruments compared to a mean for oven-drying of 13.13%. The reproducibility was 0.16%.

19.1.1.3 Particle Size

The mean particle size (MPS) is an important quality parameter for biscuit (cookie) flours. Because particle size has a pronounced effect on $\log(1/R)$ values, it follows that it can be measured by NIR. Only a single, preferably neutral, wavelength is required [3]. The relationship between MPS and $\log(1/R)$ at 2230 nm is nonlinear [3], but over the narrow range typical of U.K. biscuit flours, a sufficiently linear calibration may be obtained. An SEP of 3% was achieved over the range 57 to 74% of flour that passes a 75 μm sieve (B. G. Osborne, unpublished results).

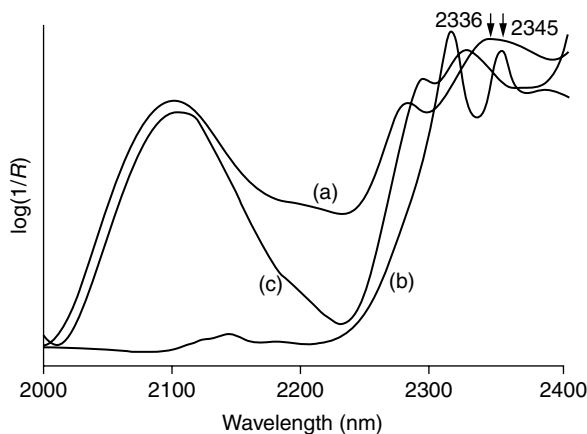


FIGURE 19.3 NIR spectra of cellulose (a), oil (b), and starch (c).

19.1.1.4 Ash

The conventional test for the purity of flour, with respect to bran contamination, is ash content. Ash is defined as the oxidized residue remaining after flour is incinerated at high temperatures. Because the ash content of bran (up to 8%) is considerably higher than that of the endosperm (0.3%), the ash content of flour is a sensitive index of bran contamination. High ash content of white flour adversely affects both its baking performance and the crumb color of bread baked from it. Therefore, flour specifications usually aim for the minimum ash content.

The incinerated residue from the ash test has no characteristic NIR absorption bands. However, one of the major components of bran is cellulose, which has a characteristic band at 2336 nm overlapping with a band due to starch (Figure 19.3). Therefore, NIR measurements at this wavelength should be an index of bran content of flour and correlate highly with ash. Empirical calibrations for ash have been reported [2,4,7] that on superficial inspection appear to perform satisfactorily. However, bran also contains higher amounts, than endosperm, of oil and protein. Consequently, ash calibrations usually contain wavelengths characteristic of oil and protein. For a given grist, ash may also be related to particle size because soft wheats tend to yield flours of higher ash content than hard wheat. The dangers of these secondary correlations are illustrated by the results of Bolling and Zwingelberg [10], who reported that, when the grist changed, the NIR “ash” followed the protein upward, although the actual ash was unchanged. Similarly, both Williams et al. [1] and Iwamoto et al. [11] reported that ash calibrations employed wavelengths associated with oil absorptions and particle size; use of these calibrations on independent samples resulted in serious errors. Williams et al. [1] proposed the incorporation of 2345 nm in ash calibrations on the grounds that it is less prone to interference from oil. However, Iwamoto et al. [11] showed that the absorption band at 2345 nm was absent in the spectrum of defatted flour and concluded that it was not associated with cellulose but a solvent-extractable substance. Their conclusion is supported by the spectra of cellulose and oil shown in Figure 19.3.

It is apparent from the results of considerable research that it is inappropriate to attempt NIR calibration against ash yield of flour and that an index such as oil, which has NIR absorption bands, might be employed. Bertrand et al. [12] proposed using oil as a marker for bran in semolina without calibration against a reference method. Their measurement wavelength of 1724 nm was also noted by Iwamoto et al. [11] as an oil band that is related to flour purity.

19.1.1.5 Color

The alternative to ash as an index of flour purity is a color measurement, because the color of the flour is directly related to the crumb color of the loaf baked from it. At least one NIR instrument

contains a filter in the visible region at (540 nm), and this allows color to be measured at the same time as protein and moisture [2]. Colour has also been used as a basis to predict ash [13]. The reflectance at 540 nm is calibrated against the AACC Agtron method, which is also a dry color measurement. However, the color of flour is a complex phenomenon that includes contributions from the bran content (grade), degree of yellowness (Agtron), and particle size. In an attempt to measure the bran content independently of the other factors, Kent-Jones and Martin designed a color grader that makes a measurement at a similar wavelength to the Agtron method, but on a slurry of flour and water. This measurement is known as the grade color figure (GCF). Reflectance measurements made at 540 nm by the NIR instrument are highly correlated with Hunterlab values for dry flour ($r = 0.95$), but only poorly correlated with GCF. However, if two wavelengths (540 and 1680 nm) are used for GCF, the correlation coefficient improves to 0.93 and the equation gives an SEP of 0.46 GCF unit [3].

19.1.1.6 Starch Damage and Water Absorption

The proportion of starch granules mechanically damaged during the milling process is an important quality factor in bread making, because the baker requires consistent flour water absorption from one batch to another. Although starch damage is not the only factor that influences water absorption, it is one that the miller can control in a given grist by adjusting the milling conditions. An NIR starch damage calibration was derived by Osborne and Douglas using a scanning monochromator [14]; then appropriate filters were obtained for this measurement to be performed on a fixed filter instrument [3]. The SEP for this calibration was 3.2 Farrand units, which is large in comparison with the range normally encountered in bread flours (20 to 40 Farrand units). A direct NIR calibration for water absorption with an SEP of 1.5% has been derived [5]. The wavelengths involved in this calibration correspond to protein, moisture, and starch damage, the major factors known to influence water absorption. However, because damaged starch plays a significant role in the Farrand equation to predict water absorption from wet chemical data, NIR must measure damaged starch reliably, in addition to the established measurements of protein and moisture, if it is to measure water absorption successfully.

Research using sets of Australian and New Zealand commercial flour samples augmented with samples produced on a 650 kg/h pilot mill resulted in a calibration with an $r^2 = .940$ and SEC $V = 0.41\%$ (Figure 19.4) [15]. In this study, the improvement to the accuracy of starch damage calibration depended on a more precise reference method than that used previously; the use of a research monochromator instrument (NIRSystems Model 5000) and its associated software (specifically, Modified Partial Least Squares regression); and careful control of sample temperature. In particular,

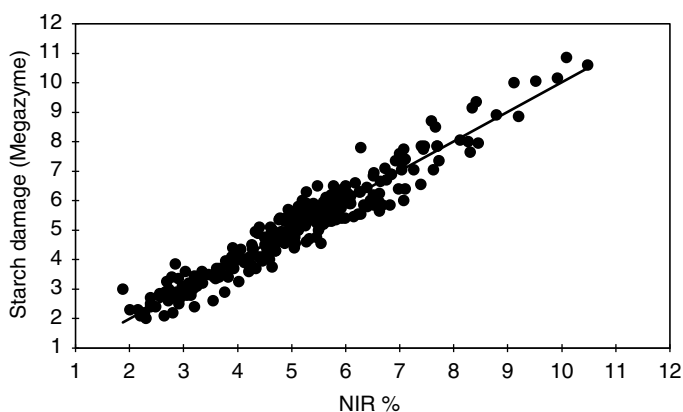


FIGURE 19.4 Plot of Megazyme vs. NIR flour starch damage.

the effect of variation in sample temperature on NIR starch damage is substantial. This means that sample temperature must either be controlled very closely or measured and a correction applied to the apparent NIR starch damage. The NIR calibration for starch damage was validated by using it to track changes that resulted from increases and decreases to reduction roll pressures during further pilot milling trials. Random samples analyzed by the reference method confirmed that the NIR and reference starch damage data follow the same trends.

19.1.2 BAKING QUALITY OF FLOUR

Because bakers are using the NIR technique for flour quality control, it is appropriate to consider whether NIR can predict the baking quality in terms of measurable properties of the bread or biscuit. The following is a summary of some unpublished results obtained by the Flour Milling and Baking Research Association, Chorleywood, from two experiments designed to explore this possibility.

Forty-nine commercially milled flours ranging from weak English to strong Canadian were baked into 400-g loaves in a carefully controlled procedure involving a constant recipe (with the exception of water absorption) and constant mixing, proving, molding, and baking conditions. NIR data on the flours were recorded using a 19-filter instrument, the protein content of the flours (8.7 to 14.5%) was determined by the Kjeldahl method ($N \times 5.7$), and the volume of the loaves (1170 to 1600 ml) was measured by the method of seed displacement. Multiple regression analysis was carried out for loaf volume vs. NIR data using all 19 $\log(1/R)$ values; then small subsets in which the standard deviation was approximately the same as for the full model were selected. Several possible subsets gave standard errors of calibration (SEC) in the range 45 to 50 ml, with protein wavelengths making the major contribution in each case; the subsets 2180, 2100, and 1680 nm gave an SEC of 50 ml. However, Kjeldahl protein figures could be used to predict loaf volume on the same samples with a SEC of 49 ml. It is clear from these results that the contribution of NIR to the prediction of loaf volume lies in its ability to measure protein content; NIR performs neither better nor worse than Kjeldahl data. Previous work with wheat carried out using a scanning monochromator [16] gave rise to the same conclusion. It was also reported [16] that NIR is unable to measure α -amylase activity. This is illustrated by the data given in Table 19.1 in which one of the 49 flours (a strong Canadian) was progressively supplemented with fungal α -amylase. This had the effect of increasing the measured loaf volume, but not the loaf volume predicted by NIR.

Over a 5-year period, a total of 75 flour samples were baked into semisweet and short-sweet biscuits using carefully controlled procedures, and NIR data on the flours were recorded using a scanning monochromator. Multiple regressions of data on four biscuit properties (water level, biscuit weight, texture, and bulk density) on the NIR data were carried out after transformation on the second derivative and with a different calibration intercept allowed for each of the

TABLE 19.1
Effect of α -Amylase Activity on Measured and NIR-Predicted Loaf Volume

Actual loaf volume (ml)	Predicted loaf volume (ml)	α -Amylase (Farrand) units
1462	1502	2
1511	1528	12
1544	1502	21
1586	1505	55
1604	1476	102

five blocks of data. Plots of correlation coefficient vs. wavelength for each of the parameters closely resembled that for flour protein, and forward stepwise regression selected known protein wavelengths.

It must be concluded that NIR is not applicable to the prediction of flour baking quality and is of limited use for measurement of ash. However, it may be used on a routine basis for measurement of protein, moisture, starch damage, water absorption, particle size, and color.

19.1.3 WHEAT GLUTEN

Gluten is separated from ground wheat and used in the form of a dried powder as a supplement in bread for two reasons: (a) to assist the base flour to carry the inert bran and sustain higher proof and loaf volumes in breads made from brown and wholemeal flours than would otherwise be possible; and (b) to replace, wholly or partly, the protein in bread flour normally derived from high-protein wheat in the grist.

NIR calibrations for protein and moisture in gluten were derived using 29 commercial samples on which protein had been determined by the Kjeldahl method ($N \times 5.7$) and moisture by oven-drying at 103°C for 16 h. NIR data were recorded on a six-filter instrument, and the statistics of the calibration equations, selected by studying all the subsets, are given in Table 19.2. Hence, NIR may be used to measure the protein content (%) of the base flour (P_F) [3] and the gluten (P_G). The amount of gluten (W_G) required to raise the protein content of the flour to P_S may be calculated from these measurements according to the formula

$$W_G = \frac{100(P_S - P_F)\%}{(P_G - P_S)}$$

Although the SEC for P_G given in Table 19.2 is large, it is negligible compared to $(P_G - P_S)$ while the error in P_F is quite large relative to $(P_S - P_F)$. Therefore, it is the error of NIR flour protein measurement that dictates the error in W_G .

19.1.4 ADDITIVE OR NUTRIENT PREMIXES

Chemical additives used in baking and not already present in flour in the required quantities are added to the dough in the form of a premix. Osborne [17] used NIR in a model experiment to determine three bread improvers (ascorbic acid, L-cysteine, and azodicarbonamide) in admixture in a starch diluent. Although the accuracy of each calibration was not satisfactory for their determination at the levels encountered in bakery premixes, the results were sufficient to indicate the potential of NIR for this type of application. This potential has been realized in the determination of nicotinamide in nutrient premixes where Osborne [18] obtained excellent results by NIR compared with high-performance liquid chromatography. The NIR calibration involved a ratio of second derivatives at 2138 nm, corresponding to a nicotinamide absorption band, and 2070 nm as a reference.

TABLE 19.2
Calibration Statistics for NIR Analysis of Gluten

Statistic	Protein	Moisture
Wavelengths (nm)	2180, 2100, 2300	1940, 2100, 2230
Range (%)	62.6 to 78.2	6.9 to 9.5
Correlation coefficient	0.980	0.985
SEC (%)	0.73	0.12

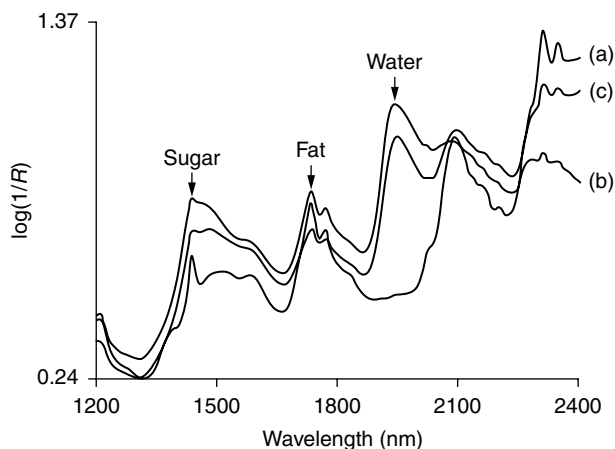


FIGURE 19.5 NIR spectra of mixtures of sugar/fat (a), sugar/fat/water (b), and sugar/fat/flour/water (c).

19.1.5 PACKAGING MATERIALS

Bakeries make extensive use of laminates for packaging of their products, and it is necessary to use the correct laminate for a particular product. A rapid method for detecting a faulty or incorrect batch of laminate is therefore desirable. In a feasibility study to test the possible application of NIR to this problem, Davies et al. [19] demonstrated that the absence of one component in a three-component laminate could be detected using a filter instrument.

19.2 DOUGH

Baking processes generally involve the blending of flour with water and other ingredients to produce a dough or batter, which is then baked in an oven to form the product. Control of the amounts of ingredients is more usefully carried out by analyzing the intermediate, as this can more easily be recycled and the wastage of energy in baking an unsatisfactory product is avoided.

NIR has been used for the compositional analysis of short-sweet biscuit doughs and dough pieces [20]. It is obvious from the spectra in Figure 19.5 that the absence of an ingredient could readily be detected from the NIR spectrum.

Quantitative calibrations, obtained using whole dough pieces, were sufficiently accurate to screen for gross errors in water, sucrose, or flour addition. The calibration for fat allowed the detection of metering errors at 5% (relative to the total amount of fat in the recipe). Although this work was carried out using a monochromator, only eight wavelengths in total were required, so that the application would be of practical value using a filter instrument.

A similar experiment has also been carried out with bread dough, and the calibration statistics are given in Table 19.3. This application was accomplished with a 19-filter instrument, and as for biscuit doughs the results indicate that NIR could be used to screen for gross errors in ingredient levels.

Dough mixing is a critical stage of the breadmaking process. During mixing, energy is imparted to the dough to develop the gluten matrix. This matrix enables the gas produced during fermentation to be retained during proving and baking. Poor mixing is one of the major, single causes of bread faults. Craft bakers judge optimum dough development by sight and touch, but in modern industrial bakeries an automated approach is required. The availability of diode array instruments, such as the Perten DA-7000, enabled the development of dynamic NIR spectroscopy and its patented application to the determination of the point of optimum dough mixing [21]. In its original implementation (Figure 19.6), the diode array instrument was placed above the mixer and reflected radiation from the moving dough mass was measured [22].

TABLE 19.3
Calibration Statistics for NIR Analysis of Bread Dough

Statistic	Protein	Fat
Wavelengths (nm)	2230, 2180, 2100, 2050	2310, 2100
Range (%)	5.2 to 7.5	0.3 to 1.6
Correlation coefficient	0.940	0.821
SEC (%), $n = 57$	0.20	0.25
SEP (%), $n = 57$	0.29	0.33



FIGURE 19.6 NIR monitoring of dough mixing.

The NIR absorption at 1160 nm decreases to the point of optimum dough mixing, as the water in the dough is bound to the protein and starch fractions, then increases as excess water is released. This provides a direct chemical measurement of the process that is independent of mixer design and speed. It has been shown to be consistent with the changes to the quality of the bread baked at different stages of mixing. Subsequently, Wesley et al. [23] employed a fibre optic interactance probe to make the same measurement in a 35 g Mixograph. The use of NIR for real-time monitoring of dough mixing has been confirmed and extended by research groups in the United Kingdom [24–27] and the United States [28–32]. These groups have shown that the NIR method works for a variety of full formula doughs in commercial-scale mixers and can be used for research into the relationships between the NIR mixing curves, flour quality, ingredient interactions, flour protein functionality, and final bread quality.

19.3 CEREAL FOODS

Bakers analyze finished products for several reasons: (a) as a check on the levels and quality of ingredients used; (b) to ensure compliance with statutory compositional requirements, for example,

TABLE 19.4
Applications of NIR to the Analysis of Cereal Products

Product	Moisture	Protein	Fat	Starch	Sugars	Fibre
Bread	✓	✓	✓	✓	✓	
Biscuits	✓		✓		✓	
Cake mixes			✓		✓	
Breakfast cereals					✓	✓
Pasta	✓	✓	✓			
Snack foods				✓		✓

TABLE 19.5
Accuracy of NIR Calibrations for Sliced Bread

Calibration	SEC (%)	SEP (%)	Reference
Protein	0.09	0.20	[33]
	0.24	—	[34]
	0.23	0.22	[35]
Fat	0.16	0.18	[33]
	0.13	—	[34]
	0.24	0.20	[35]
Moisture	0.46	0.51	[33]
	0.67	0.62	[35]
Dry matter	0.57	—	[34]

moisture content of bread; and (c) to provide data for nutrition labeling. Few bakeries have chemical laboratories; therefore a technique such as NIR is required if they are to carry out these analyses themselves, and a number of applications of NIR to the analysis of baked products have been published (Table 19.4).

19.3.1 BREAD

NIR had been employed in the United Kingdom, United States, and South Africa for the determination of protein, fat, and moisture in sliced bread [33–36]. Satisfactory calibrations were obtained by taking circular subsamples from each of 5 to 6 slices and averaging the NIR data, provided that precautions were taken to prevent moisture loss from the disks during analysis. Drying and grinding of the bread did not significantly improve the accuracy of the NIR analytical method [33,35]. The standard errors (Table 19.5) were remarkably consistent between the different studies and acceptable for routine use (coefficients of variation 1 to 2%). The calibrations also proved stable across the variability found between bakeries [36] and readily transferable between bread baked by widely different procedures and flour [35]. Calibrations have also been reported for total sugars in bread slices [35], with an SEC of 0.34% in the range 0.9 to 5.6%; and starch in air-dried bread [37], with an SEC of 2.0% in the range 51 to 65%.

If bread crumb is stored for a few hours or more, the process known as staling or retrogradation occurs. In one of the first reports of the use of NIR for the study of reaction kinetics in food systems, second derivative data at 1414 and 1465 nm were used to calculate rate constants for the retrogradation process in bread [38,39]. NIR spectra were recorded using discs sampled from bread

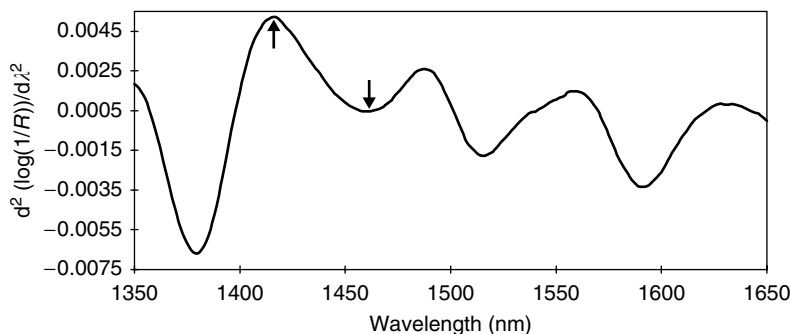


FIGURE 19.7 Difference spectrum for stored and fresh bread.

slices taken from loaves during storage. A generic spectroscopic rationale for this application was proposed by Osborne [40]: Retrogradation occurs through the slow redevelopment of crystallinity in the amylopectin fraction of the starch. This crystallinity is a direct result of extensive hydrogen bonding, both intramolecularly and to water molecules, of the amylopectin molecules. It is these changes to the hydrogen bonding of the starch and water O—H bonds at 1410 and 1460 nm that are measured by the NIR spectra of the intact bread crumb (Figure 19.7).

19.3.2 BISCUITS

The use of NIR in the monitoring of automatic metering equipment used for fat, sucrose, dry flour, and water in short dough biscuit products has been investigated [20]. This study introduced an experimental design to minimize intercorrelations between the major variables and a wavelength selection strategy that became widely used for monochromators. A standard recipe was varied to provide as great a range as possible for each of the four constituents under investigation. To prevent high correlations between fat and sucrose, the recipe was randomized by varying independently the amounts of these two ingredients. Then, for each recipe, the amounts of water and flour were calculated so that the dough produced would be capable of manipulation on a rotary molder. Whole biscuits were presented to an NIR scanning monochromator with the writing and docker pin holes facing the light beam, as would be the case on a production line. The wavelength selection strategy involved selection of wavelengths two at a time to find the best pair. Choosing two variables together in this way overcomes the multicollinearity problem, and when a two-term equation is adequate this procedure should find it. If necessary, further wavelengths may be selected by forward stepwise regression. NIR calibrations for fat, sucrose, dry flour, and water in whole biscuits were thus obtained. However, although the calibration had high correlation coefficients and acceptable precision compared with other methods, only that for fat was capable of sufficient accuracy for the detection of ingredient metering errors likely to be encountered in a commercial bakery.

19.3.3 FLOUR CONFECTIONERY

The NIR calibrations for doughs, breads, and biscuits described previously were derived by means of designed experiments carried out under carefully controlled conditions in pilot bakeries. However, NIR is used under commercial conditions, and an example is the analysis of dry cake mixes [41]. A 19-filter instrument was used for this application that was calibrated in the laboratory, and then assessed under quality-control conditions in the bakery. It was desirable to have a single calibration for a range of different mixes (bread, scone, sponge, shortbread, and shortcrust), and this was achieved for fat and sucrose. The calibration for fat gave an SEP of 1.7% in the range 3 to 26% while that for sucrose gave an SEP of 2.7% in the range 13 to 41%. Although these SEP figures are much higher

than those obtained between replicate determinations using the corresponding reference methods (fat by Soxhlet of 0.38% and sucrose by HPLC of 1.04%), they proved acceptable for quality-control purposes in view of the speed with which NIR results could be obtained.

19.3.4 BREAKFAST CEREALS

Baker et al. [42,43] investigated the relationship between NIR measurements and neutral detergent fiber (AACC method 32 to 20) values of breakfast cereals. In a preliminary study [42] a correlation coefficient of 0.985 and an SEC of 1.36% over the range 0 to 31.7% were obtained using 213 samples of various cereals based on wheat, maize, oats, and rice. These results were achieved by means of a two-term calibration equation in which the NIR data were treated as first derivative of $\log(1/R)$ divided by the first derivative of $\log(1/R)$ at a reference wavelength. Further work with a different monochromator instrument [43] resulted in a very similar equation with SEC of 1.5%; testing this equation on independent samples gave an SEP of 1.8%. Calibration wavelengths for fiber (1352, 2204, 2448, and 2450 nm) had to be selected so as to avoid interference caused by the presence of crystalline sucrose in the samples. Because of the interference, it became apparent that it might be possible to determine the total sugar content of breakfast cereals by NIR. In an experiment designed to test this hypothesis [44], 84 branded cereals, of which the total sugar content ranged from 0 to 54%, were used. These were divided into calibration and prediction sets of 40 and 44 samples, respectively, and a calibration equation derived using first-derivative ratios of 1432/1822 and 1692/1782 nm. This equation gave an SEC of 2.7% on the calibration samples and an SEP of 2.8% on the prediction samples against the reference gas-liquid chromatography method. The SEP is in very close agreement with that reported for sucrose in cake mixes [41] and confirms that NIR is applicable to the determination of sugars in cereal foods.

19.3.5 PASTA

Kaffka et al. in two carefully controlled experiments [45,46], studied the application of NIR to the determination of fat, protein, moisture, and egg contents of ground pasta. In the first of these experiments [45], pasta samples were prepared with 0, 2, 4, or 6 whole eggs per kilogram, and fat and protein were determined by chemical analysis. Although excellent calibrations were achieved, the same wavelengths were involved in each and it was concluded that the egg, fat, and protein contents were highly intercorrelated. Although fat and protein may be suitable markers from which to measure the egg content of pasta, these two constituents should be determined independently of each other if the application is to be used successfully. In the second experiment [46], therefore, the pastas were prepared by adding egg yolk and albumin in varying proportions in an attempt to randomize the fat and protein levels; the actual correlation between the levels of these constituents was 0.362, but this was reduced further by incorporating samples having the same composition but ground to different particle sizes. Single-term calibrations involving ratios of second derivatives of $\log(1/R)$ data were derived for fat, protein, and moisture with SEC figures at 0.11, 0.24, and 0.24%, respectively.

19.3.6 EXTRUSION COOKING

Extrusion cooking is used to manufacture from cereals a wide range of products including breakfast cereals, snack foods and pet foods. In these products, texture and density are key quality requirements that are controlled by the processing conditions summarized by the term "degree of cook." Conventionally, control is attempted by measurement of inputs such as the moisture content, screw speed, barrel temperature and so forth on the basis that the relationship between these and the product quality is understood. However, a more direct means of control is highly desirable.

In contrast to breadmaking, where the product structure is largely controlled by protein development, in extrusion cooking it is the starch component of the cereal-based raw material that is the chief functional ingredient. After crystalline regions of starch granules melt at high temperatures and pressures, the application of powerful shearing forces in the extruder soften granular structures and cause disruption and dispersal. A continuous phase of starch polymers containing bubbles of steam in the extrudate allows for an extensible foam-like structure to develop. The extent of transformation of the raw material, referred to as the “degree of cook,” is crucial to final product quality. Degree of cook increases when there is an increase in depolymerization of the starch molecules, which in turn results in an increase in the number of free hydroxyl bonds and a decrease in paste viscosity. The degree of cook of a powdered extrudate sample may be measured by means of a viscometer. However, viscometry can only be carried out on a sample that has already been processed, and hence is not suitable for feedback process control.

A project with the aim of using NIR spectroscopy online for the measurement of degree of cook in real time was carried out within the EU Fourth Framework R&D Programme by UK and French research laboratories, a Swedish instrument manufacturer, and a Swiss snack food company. Initially, products from wheat, maize, oat, and rice were prepared over a wide range of processing conditions and the extrudates freeze-dried and ground to powders before NIR spectroscopy. Successful calibrations based on spectral features related to starch structure were obtained using the powdered samples [47,48] and on-line [49]. A fibre optic reflectance probe was installed in the die of the extruder and the light was transmitted through the melt and reflected back from a polished steel pillar. The spectral characteristics of the different forms of starch found in the hot melts close to the die of the extruder were identical to those of the powdered extrudates [49]. In Australia, online NIR transmission spectroscopy using paired fiber optic probes was used to monitor the physical state of starch polymers in the barrel melt as a means of measuring the degree of change that has occurred under a given processing regime [50]. An NIR measurement based on the band at 1430 nm identified as a result of the strategic research on wheat starch structure [40] was used to monitor changes that resulted from adjustments to different process variables.

19.3.7 MISCELLANEOUS CEREAL FOODS

The applications of NIR to the analysis of baked products discussed so far were designed for specific classes of samples such as bread or biscuits, although in some cases [41–44] calibrations have been derived for a range of individual products of the same class. The use of sample-specific calibrations in a bakery, however, is highly inconvenient; because wide ranges of products need to be analyzed, many calibrations would be required. Apart from the problem of exceeding the memory capacity of the simpler instruments, this introduces a greater likelihood of systematic error due to the wrong calibration being employed. NIR researchers are beginning to realize the possibility and value of more universal calibrations.

Baker [51] used NIR to determine fiber and starch in a wide range of ground snack foods (potato chips, corn chips, extruded snacks, popcorn, crackers, and pretzels) with SEP values of 1.4 and 3.5%, respectively. Ratios of derivatives of $\log(1/R)$ at wavelengths characteristic of the constituent under study and of interferences were used in these calibrations. Osborne [52] derived a single calibration for the determination of fat in bread, biscuits, and breakfast cereals. This calibration using $\log(1/R)$ data of 1720 and 1700 nm after normalization by the MSC algorithm gave an SEP of 1.10% over a range of fat contents 1.2 to 27.3%. It was, however, necessary to freeze-dry the bread and to grind all the samples before NIR analysis. Therefore, although a universal calibration could be obtained, more sample preparation was required. Furthermore, while the SEP obtained was better than that for cake mixes, it was considerably worse than that for bread (Table 19.5). Kays and Barton II reported calibrations for dietary fibre [53], protein [54], and energy content [55] in a wide range of dry-milled cereal products.

19.4 CONCLUSION

NIR is widely applicable to the banking industry for the analysis of ingredients, process intermediates, and final products, and a number of successful calibrations have been described in the literature and discussed in this chapter.

REFERENCES

1. P. C. Williams, B. N. Thompson, D. Wetzel, et al. *Cereal Fds. Wld.*, 26: 234 (1981).
2. V. R. Diachuk, E. Hamilton, N. Savchuk, and S. S. Jackel, *Bakers Dig.*, 55: 72 (1981).
3. B. G. Osborne, S. Douglas, and T. Fearn, *J. Food Technol.*, 17: 355 (1982).
4. H. Bolling and H. Zwingelberg, *Getreide, Mehl Brot*, 36: 197 (1982).
5. B. G. Osborne, S. Douglas, and T. Fearn, in *Progress in Cereal Chemistry and Technology* (J. Holas and J. Kratochvil, eds.), Elsevier, Amsterdam (1983), Part A, pp. 577–581.
6. B. G. Osborne and T. Fearn, *J. Fd. Technol.*, 18: 453 (1983).
7. M. Iwamoto, C. R. Kwang, T. Suzuki, and J. Uozumi, *Nippon Shokuhin Kogyo Gakkaishi*, 31: 50 (1984).
8. B. G. Osborne and G. M. Barrett, *Milling*, 179: 26 (1986).
9. B. G. Osborne, *J. Sci. Fd. Agric.*, 38: 341 (1987).
10. H. Bolling and H. Zwingelberg, *Getreide, Mehl Brot*, 38: 3 (1984).
11. M. Iwamoto, N. Kongsree, J. Uozumi, and T. Suzuki, *Nippon Shokuhin Kogyo Gakkaishi*, 33: 842 (1986).
12. D. Bertrand, P. Robert, M. F. Devaux, and J. Abecassis, in *Analytical Applications of Spectroscopy* (C. Creaser and A. M. C. Davies, eds.), Royal Society of Chemistry, London (1988), pp. 450–456.
13. D. S. Reed and J. J. Psotka, US Patent 5,258,825 (1991).
14. B. G. Osborne and S. Douglas, *J. Sci. Fd. Agric.*, 32: 328 (1981).
15. B. G. Osborne, *Proc. Australian Cereal Chemistry Conference* (L. O'Brien, A. B. Blakeney, A. S. Ross, and C. W. Wrigley, eds.), RACI, Melbourne (1998), pp. 434–438.
16. B. G. Osborne, *J. Sci. Fd. Agric.*, 35: 106 (1984).
17. B. G. Osborne, *J. Sci. Fd. Agric.*, 34: 1297 (1983).
18. B. G. Osborne, *Analyst, Lond.*, 112: 313 (1987).
19. A. M. C. Davies, A. Grant, G. M. Gavrel, and R. V. Steeper, *Analyst, Lond.*, 110: 643 (1985).
20. B. G. Osborne, T. Fearn, A. R. Miller, and S. Douglas, *J. Sci. Fd. Agric.*, 35: 99 (1984).
21. I. J. Wesley, N. Larsen, B. G. Osborne, and J. H. Skerritt, US Patent 6,342,259 B1 (2002).
22. J. Wesley, N. Larsen, B. G. Osborne, and J. H. Skerritt, *J. Cereal Sci.*, 27: 61–69 (1998).
23. J. Wesley, O. Larroque, B. G. Osborne, N. Azudin, H. Allen, and J. H. Skerritt, *J. Cereal Sci.*, 34: 125–133 (2001).
24. J. M. Alava, S. J. Millar, and S. E. Salmon, *J. Cereal Sci.*, 33: 71–81 (2001).
25. J. M. Alava, S. J. Millar, and S. E. Salmon, *Spec. Publ. R. Soc. Chem.*, 261: 439–441 (2000).
26. S. J. Millar and J. M. Alava, Paper 39, 87th AACC Annual Meeting; www.aaccnet.org/meetings/2002/abstracts (2002).
27. S. J. Millar, J. M. Alava, and S. E. Salmon, Paper 186 AACC Annual Meeting, October 14–18, Charlotte, NC; www.aaccnet.org/meetings/2001/abstracts (2001).
28. R. E. Dempster, M. C. Olewnik, and V. W. Smail, AACC Annual Meeting, Portland, Paper No. 39 (2003).
29. W. N. Huang, Paper 28 AACC Annual Meeting, October 14–18, Charlotte, NC; www.aaccnet.org/meetings/2001/abstracts (2001).
30. W. N. Huang, M. C. Olewnik, J. J. Psotka, and R. E. Dempster, Paper 133 AACC Annual Meeting, October 14–18, Charlotte, NC; www.aaccnet.org/meetings/2001/abstracts (2001).
31. J. Psotka and M. Olewnik, Paper 303 AACC Annual Meeting, November 5–9, Kansas City, MO (2000).
32. J. Psotka, R. Chen, and M. Olewnik, Paper 293 AACC Annual Meeting, October 31 to November 3, Seattle, WA (1999).
33. B. G. Osborne, G. M. Barrett, S. P. Cauvain, and T. Fearn, *J. Sci. Food Agric.*, 35: 940 (1984).
34. C. A. Groenewald, *Supplement to SA Food Review*, 59 (1984).

35. K. Suzuki, C. E. McDonald, and B. L. D'Appolonia, *Cereal Chem.*, **63**: 320 (1986).
36. E. Duvenage, *J. Sci. Food Agric.*, **37**: 384 (1986).
37. B. G. Osborne, *Anal. Proc.*, **20**: 79 (1983).
38. R. H. Wilson, B. J. Goodfellow, P. S. Belton, B. G. Osborne, G. Oliver, and P. L. Russell, *J. Sci. Food Agric.*, **54**: 471–483 (1991).
39. B. G. Osborne, *Analysis*, **26**: M39–M41 (1998).
40. B. G. Osborne, *J. Near Infrared Spectrosc.*, **4**: 195–200 (1996).
41. B. G. Osborne, T. Fearn, and P. G. Randall, *J. Food Technol.*, **18**: 651 (1983).
42. D. Baker, K. H. Norris, and B. W. Li, in *Dietary Fibers, Chemistry and Nutrition* (G. E. Inglett and S. L. Falkenhag, eds.), Academic Press, New York (1979), pp. 72–78.
43. D. Baker, *Cereal Chem.*, **60**: 217 (1983).
44. D. Baker and K. H. Norris, *Appl. Spectrosc.*, **39**: 618 (1985).
45. K. J. Kaffka and F. Kulcsar, *Acta Aliment.*, **11**: 47 (1982).
46. K. J. Kaffka, K. H. Norris, and M. Rosza-Kiss, *Acta Aliment.*, **11**: 199 (1982).
47. R. C. E. Guy, B. G. Osborne, and P. Robert, *J. Food Eng.*, **27**: 241–258 (1996).
48. S. Millar, P. Robert, M. F. Devaux, R. C. E. Guy, and P. Maris, *Appl. Spectrosc.*, **50**: 1134–1139 (1996).
49. R. C. E. Guy, *Extrusion Communiqué*, May–June, 18–19 (1996).
50. A. J. Evans, S. Huang, B. G. Osborne, Z. Kotwal, and I. J. Wesley, *J. Near Infrared Spectrosc.*, **7**: 77–84 (1999).
51. D. Baker, *Cereal Food World*, **30**: 389 (1985).
52. B. G. Osborne, in *Analytical Applications of Spectroscopy* (C. Creaser and A. M. C. Davies, eds.), Royal Society of Chemistry, London (1988), pp. 68–71.
53. S. E. Kays and F. E. Barton II, *J. Near Infrared Spectrosc.*, **6**: 221–227 (1998).
54. S. E. Kays and F. E. Barton II, *J. Near Infrared Spectrosc.*, **8**: 35–43 (2000).
55. S. E. Kays and F. E. Barton II, *Near Infrared Spectroscopy: Proceedings of the 10th International Conference* (A. M. C. Davies and R. K. Cho, eds.), NIR Publications, Chichester, UK, pp. 217–220 (2002).

20 NIR Analysis of Dairy Products

Rob Frankhuizen

CONTENTS

20.1	Introduction	415
20.1.1	Use of NIRS in the Dairy Industry	416
20.1.2	Standard Methods	417
20.2	Calibration Procedure	417
20.3	Milk	418
20.4	Milk Powders	420
20.4.1	General	420
20.4.2	Results and Discussion	421
20.4.2.1	Protein	421
20.4.2.2	Moisture	422
20.4.2.3	Fat	424
20.4.2.4	Lactose	424
20.4.2.5	Lactate	425
20.4.2.6	Ash	425
20.4.2.7	Skim Milk Powder Content	426
20.4.3	Repeatability and Reproducibility	426
20.5	Casein and Caseinates	427
20.6	Butter	428
20.7	Cheese	429
20.7.1	Sample Preparation	430
20.7.1.1	Effect on Precision	430
20.7.1.2	Alternative Sample Preparation	430
20.7.2	Results: Major Constituents	431
20.7.3	Results: Minor Constituents	431
20.7.4	Determination of Ripening Time	432
20.7.4.1	General	432
20.7.4.2	Results	433
20.8	Conclusion	434
	Suggested Readings	435
	References	436

20.1 INTRODUCTION

Economic and legal constraints require strict control of the composition of dairy products. For example, the level of butterfat, which is the most valuable ingredient in raw milk, is controlled by law or by customers' specifications in a whole range of dairy products. In the Netherlands the quality

payment scheme to farmers, which applies differential payments depending among others on fat and protein level, forces the industry to test many thousands of raw milk samples. On the other hand, the worldwide market is becoming highly competitive and there often is a great surplus of powder, butter, and cheese, particularly in Europe and New Zealand. The introduction of milk quotas means farmers will need rations even more finely tuned to produce the right levels of butterfat, protein, and casein. The industry also supplies milk-based baby foods and other dietary products. It is essential that, for nutritive reasons, these products comply with the claimed composition.

The official reference analytical procedures for moisture, fat, protein, and lactose all take too long to be satisfactory for controlling a manufacturing process. Consequently, there is an urgent need for instruments that can rapidly determine quality or other important economic parameters [1–3].

Near-infrared (NIR) reflectance spectroscopy (NIRS), in particular, potentially provides these benefits and has been practiced now for some years [4].

Rapid analysis of milk for fat, protein, and total solids content by infrared (IR) absorption spectroscopy already has had a significant impact on the dairy industry [5] and is an approved AOAC method [6]. However, IR absorption spectroscopy is only sui for liquid samples; consequently, milk powder, cheese, and other solid products have to be blended and homogenized by dissolving before analysis [7]. Reflectance spectroscopy measures the intensity of light reflected from the surface of a sample, thereby eliminating the need for this sample preparation. In particular, NIR spectroscopy NIRS makes use of this principle. NIR spectra for casein, butter, and dried milk have been determined by Goulden [8], resulting in the use of several wavelengths to determine sample composition.

The application of NIR to determine the chemical composition of milk powders has been thoroughly investigated [9–13]. Birth and Washam improved the reproducibility of reflectance measurement on blue cheese [14]. Giangiacomo et al. reported significant correlations between reflectance measurements and composition of powdered, freeze-dried blue cheese [15]. Data of Frank and Birth indicate extensive overlapping between NIR absorption bands of major milk constituents [16].

This is typical for NIR spectra of agricultural products. Absorbance measurements at three to eight wavelengths are usually necessary to determine adequately the concentration of a particular constituent. Over the past few years the State Institute for Quality Control of Agricultural Products (RIKILT) in the Netherlands investigated the suitability of NIRS for measuring the composition of some dairy products such as raw milk, butter, casein and caseinate, milk powders, and cheese. Correlations between major and some minor constituents analyzed by classical chemical methods and NIRS reflectance measurements were determined. For this wavelengths were computer-selected and unsuitable wavelengths were eliminated, based on statistical parameters and product or constituent information [17–21].

20.1.1 USE OF NIRS IN THE DAIRY INDUSTRY

Near-infrared spectroscopy can be used in different areas of the dairy industry. It is possible to survey applications in three ways:

1. Analysis of the incoming milk in terms of fat, protein, and lactose for payments to farmers and standardization of the milk in terms of fat, protein, and solids.
2. Analysis of the (spray)-dried products on fat, moisture, and protein to get the right composition mixture for special products and analysis of the raw materials for the production of cream, cheese, casein, and whey.
3. Analysis of the finished products to control the products in terms of legal regulations and customers' or producers' specifications.

Governmental or other official laboratories for quality control of dairy products can use NIRS as a screening method for rapid analysis of the composition of unknown samples. Besides with NIRS

it might be possible to determine quality parameters where in the past no direct correlation with chemical or physical methods was apparent.

20.1.2 STANDARD METHODS

The standard methods of laboratory analysis by wet chemistry vary from country to country. However, the following methods are generally accepted as standard methods in the dairy industry [22]:

Component	Method
Moisture	Toluene
	Oven
	Karl Fisher
	Vacuum
Fat	Gerber
	Soxhlet extraction
	Gravimetric
	Röse–Gottlieb
	Schmid–Bondzynski–Ratzlaff extraction
	Mojonnier
Protein	Babcock
	Kjeldahl
	Dye binding
Lactose	Photometric
	Reductometric
	Polarimetric
	Titrimetric
	Enzymatic
Lactic acid and lactates	Photometric
	Enzymatic
Ash	Furnace
Salt	Potentiometric
	Titrimetric
	Gravimetric
pH	Electrometric

20.2 CALIBRATION PROCEDURE

Near-infrared spectroscopy measurements are as good as the calibration. That is why great accuracy and expense has to be made by calibrating a NIRS instrument. Calibration is achieved by taking a set of samples that have been analyzed by reference methods for parameters of interest. A minimum of 40 calibration samples normally is required in order to adequately calibrate the instrument. It is very important that a set of calibration samples represents the full range of the product being analyzed. For achieving a good “robust” calibration it is therefore necessary that samples are collected over a period of time depending upon composition, production process, production period, and production area. Definite minimum and maximum ranges are difficult to set, however, as a general guideline the range should not go smaller than 2% and not exceed 20% for major constituents. It is important to know that most dairy products will change in composition over a period of time. Milk powder absorbs and dissipates moisture very readily. Products containing high levels of fat are liable to fat oxidation over

a period of a few months and products containing high levels of lactose or other carbohydrates may undergo some form of crystallization. The oxidation and crystallization processes can change the absorbance characteristics of fat and carbohydrates. Adequate storage will prevent significant changes in absorbance characteristics, however, it is better to use fresh samples and analyze the samples with the reference methods and NIRS at the same time. Accurate and consistent laboratory analysis is important in order to obtain the best results with the NIR technique. When sufficient samples with the required range are collected a regression analysis between the analytical data and the optical density values of the samples can be carried out and will result in a set of constants (K values). Each product and constituent has its own set of constants. Statistical tests, incorporated in the software of the instrument, enables the user to choose the proper filter or wavelength combination for his calibration. Absorbance measurements at 3 to 8 wavelengths are usually necessary to determine adequately the concentration of a particular constituent. Standard deviation, correlation coefficient and the t -test are the most important parameters for achieving a good calibration by using a multiple linear regression program. The ultimate test of any calibration is to check the validity of results obtained on samples that did not belong to the calibration set. Assuming a good conformity with the reference methods the calibration can be used for routine measurements. It is important that the validity of the calibration is checked frequently. (Sample selection/handling and data analysis will be discussed in more detail in Chapter 13.)

20.3 MILK

Milk has an NIR absorption spectrum very similar to that of water [23] (Figure 20.1). The difference between milk and water is that milk contains scattering particles in the form of fat globules and protein micelles. A beam of electromagnetic irradiation traveling through such a light-scattering medium takes a path longer than the sample thickness [24], with a subsequent increase in absorption, or intensification of the absorption bands [10,25]. The introduction of light particles can result in erroneous information due to the wavelength dependence of the extinction [26] and an untrue, partial collection of the transmitted light [27]. These problems can be overcome by using an integrating sphere to collect the transmitted light [25].

Another problem is the very broad and intensive absorption bands of water that overlap the absorption bands of protein and fat in the NIR milk spectrum for a great part (Figure 20.1). This overlap makes it difficult to correlate specific bands to the constituents from which they arise. Besides, the small variation of protein and fat content in standardized milk may cause only very slight changes in the NIR spectrum. Therefore, it is necessary to transform the raw data to first and second derivatives to interpret and correlate the overlapping absorption bands of protein and fat in

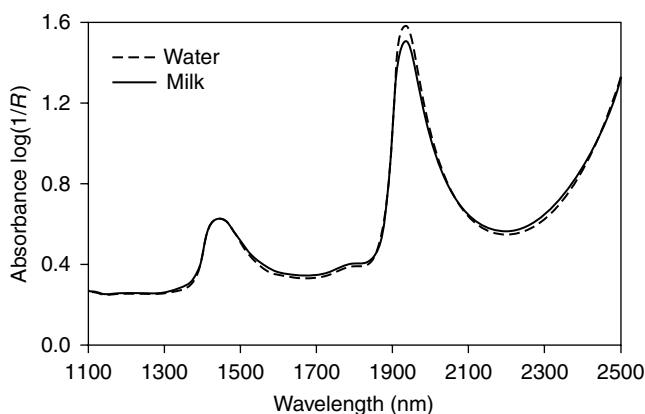


FIGURE 20.1 NIR reflectance spectra of milk and water.

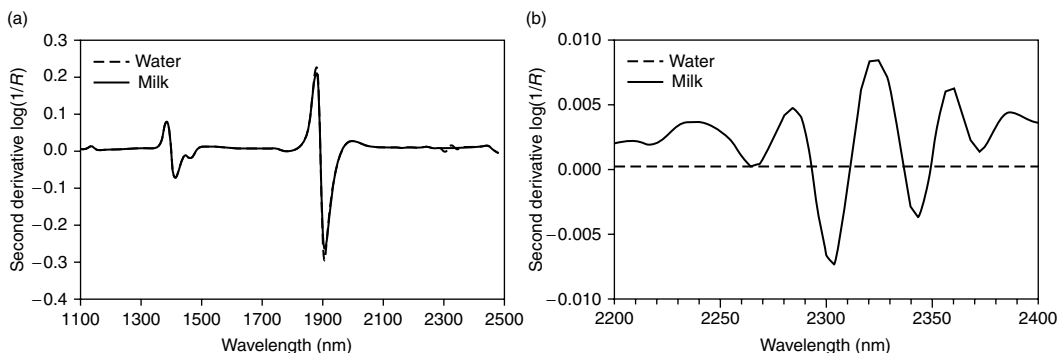


FIGURE 20.2 (a) Second derivative NIR spectra of milk and water. (b) Second derivative NIR spectra of milk and water with specific absorption bands of fat at 2304 and 2343 nm, respectively.

TABLE 20.1

NIRS Data (g/100 g) for Raw Cow Milk (Mixed Samples), Measured with a InfraAlyzer 500 Equipped with a Liquid Sample Cell and Homogenizer, Compared to Reference Methods

Constituent	<i>n</i>	SEP	Range	Accuracy required
Fat	50	0.034	4.1–7.5	0.04
Protein	50	0.038	2.1–2.8	0.04
Lactose	50	0.043	3.5–4.1	0.05

n = number of samples; SEP = standard error of prediction; reference method fat: Gerber; protein: Kjeldahl; lactose: enzymatic.

the NIR milk spectrum (Figure 20.2a and Figure 20.2b). In this way reliable and accurate results can be obtained.

Milk and its derivatives can be analyzed by NIRS for fat, protein, lactose, and total solids content using a special drawer unit equipped with a liquid cell. It uses the principle of transreflectance, which combines the advantages of transmission and of reflectance. The reflected light measured in this case originates both from the surface of the liquid and from the rear of the cell. In the latter case it has passed through the thin film of liquid contained in the cell. A high-pressure pump and a two-stage homogenization will homogenize the milk emulsion to particle size well below a micrometer. The unit also provides for temperature control and automatic wash from the NIR instrument. For achieving an accurate measurement it is absolutely necessary that the samples be freshly analyzed after warming to about 45° in a waterbath and mixing with a mixer at the inlet of the homogenizer. Study of the suitability of NIRS for the determination of fat, protein, and lactose of milk (mixed samples of several individual cows) for payments and standardization purposes shows that the accuracy of NIRS is acceptable (Table 20.1).

Study of the usefulness of NIRS for the determination of fat, protein, and lactose of goat's milk shows that the standard errors of calibrations (SECs) of a test set originating from 30 individual goats (Table 20.2) is about five to ten times higher than the SECs of the separate calibration based on mixed samples (Table 20.3). This is caused not only by an increased variation in the sample composition but especially by deviation of the fat and protein composition caused by udder disease (mastitis) of some goats. However, by addition of the individual samples, including the deviating samples, to the calibration set and by an increase of the number of wavelengths, it was possible to calculate a calibration with more acceptable accuracy (Table 20.4).

TABLE 20.2
NIRS Calibration Data (g/100 g) for Raw
Goat's Milk (Individual Goat Samples)
Compared to Reference Methods

Constituent	<i>n</i>	SEC	Range
Fat	30	0.226	2.8–7.0
Protein	30	0.170	2.8–4.1
Lactose	30	0.281	3.3–5.1

TABLE 20.3
NIRS Calibration Data (g/100 g) for Raw Goat's
Milk (Mixed Samples of Several Individual
Goats) Compared to Reference Methods

Constituent	<i>n</i>	SEC	Range
Fat	20	0.022	3.0–6.3
Protein	20	0.028	3.0–3.9
Lactose	20	0.058	3.5–4.9

TABLE 20.4
NIRS Calibration Data (g/100 g) for Raw
Goat's Milk (Mixed and Individual Samples)
Compared to Reference Methods

Constituent	<i>n</i>	SEC	Range
Fat	50	0.096	2.8–7.0
Protein	50	0.079	2.8–4.1
Lactose	50	0.133	3.3–5.1

It must be concluded that the measurement of fat, protein, and lactose of milk by NIRS can be successful. Our experience with cow and goat milk shows that the main problem is to find a representative sample set, a method of presenting the sample to the instrument that is exactly reproducible, and a method to identify unknown samples with an abnormal spectra. Study of Bertrand et al. [28] illustrates that principal components analysis (PCA) has the potential to select the most relevant calibration samples (independent of the chemical composition) and to identify samples with abnormal spectra as an outlier.

20.4 MILK POWDERS

20.4.1 GENERAL

Milk powder has been found to be one of the most suitable products to be analyzed using NIRS because of its uniform particle size and shape as well as its consistent formulation. Because milk

powder has a small particle size, it doesn't require grinding. It is important, however, that the samples be well mixed and blended. Powder is simply inserted into a sample cup, leveled off, and sealed with a backing plate. For achieving good "robust" calibrations we collected milk powders from the most important milk powder-producing EEC countries during a period of 2 years. Depending on composition, production process, and production period, samples were selected for the calibrations. The samples were analyzed in duplicate according to the following methods: Moisture was determined by oven drying (2 h at 102°C), protein (total nitrogen \times 6.38) by the Kjeldahl method, and fat by the Rose–Gottlieb extraction method. Ash content was determined by dry ashing (1 h at 500°C) and enzymatic methods were used to measure lactose and lactate content. The mean value for the skim milk powder content in calf-feed was calculated by using results from an enzymatic coagulation of the paracasein determined according to EEC method L296/10 [29]. Reflectance measurements were obtained by using an InfraAlyzer 400 after the powder had reached room temperature.

All combinations were tested by selected samples representative of the calibration. During the test period it was not always possible to get a test set with the same range as the calibration set. Therefore the *R* values of the test sets are sometimes smaller than those of the calibration sets.

To ensure that the correct wavelength was chosen, a spectral scan of a few different milk powders and of its isolated constituents protein, fat, and lactose was carried out using an InfraAlyzer 500 (Figure 20.3 and Figure 20.4). The most useful interference filters for analyzing fat, protein, moisture, and lactose in milk powders are shown in Figure 20.4 and given in Table 20.5.

20.4.2 RESULTS AND DISCUSSION

20.4.2.1 Protein

High correlations were observed between the protein percentages determined by the Kjeldahl method and those estimated by the InfraAlyzer 400 (Table 20.6). The square multiple correlation coefficients of 0.97 to 0.999 with SEC of 0.18 to 0.34% are acceptable compared with the standard deviation of the reference method of about 0.15% for milk powders.

Furthermore, it is important to realize that the Kjeldahl method determines the crude instead of the true protein percentage. The crude protein percentage, however, can differ more or less from the true protein percentage, depending on the season and resulting in fluctuations in the percentage of nonprotein nitrogen (NPN). Probably NPN differs in the NIR absorption from the true protein, likely resulting in an increased SEC compared with a calibration against the true protein percentages.

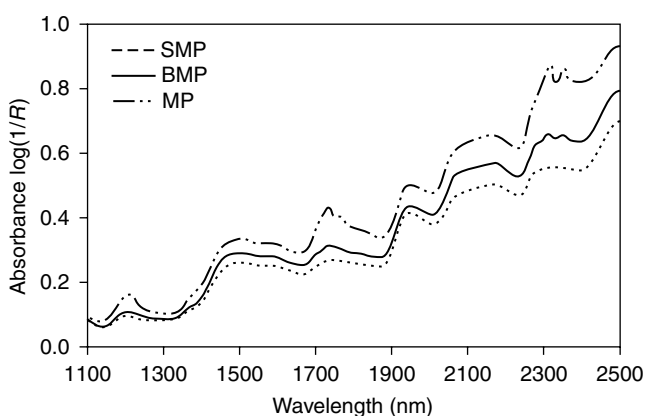


FIGURE 20.3 Typical NIR reflectance spectra of skim milk powder, buttermilk powder, and full cream milk powder.

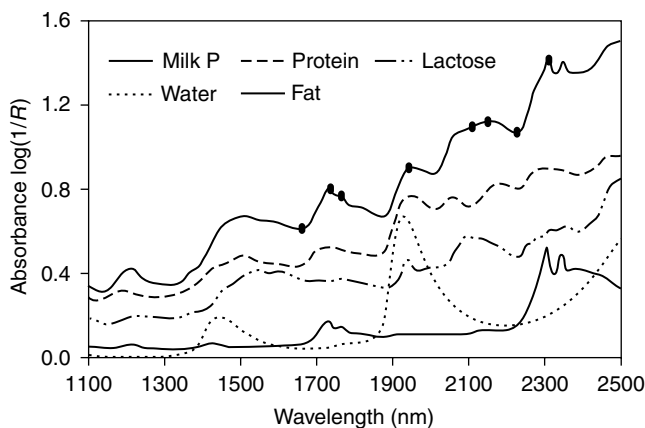


FIGURE 20.4 NIR reflectance spectra of skim milk powder and various isolated constituents (=selected wavelength).

TABLE 20.5

Absorption Bands of Milk Powder Constituents

Filter no. ^a	Absorption band (nm) ^b	Constituents contributing to absorption
20	1650	Water, protein, lactose reference
16	1940	Water
17	1734	Protein, fat
15	1759	Fat, protein
14	2100	Protein, lactose
9	2139	Protein, lactose, water reference
6	2230	Fat, protein, lactose reference
4	2310	Fat, protein

^a Filters of the Technicon InfraAlyzer 400 that correspond to the given wavelength bands.

^b Approximate wavelength.

Concerning the prediction sets (test sets), the results for the SEP and *R* values are comparable with those of the calibration sets. It can be concluded that the calibration are robust and reliable. Calibration procedures resulted in four to eight filters giving significant information, including filters for protein absorption at 2139 and 1734 nm.

20.4.2.2 Moisture

Correlation coefficients were calculated between moisture determined by oven drying and moisture estimated by the InfraAlyzer 400 (*R* = .97, tot. 0.99). Averaged SECs of 0.12% are acceptable (Table 20.7). The results obtained for the prediction sets are comparable with those of the calibration sets. The calibration procedure resulted in three to six filters giving significant information including the most significant filter for moisture absorption at 1940 nm.

Moisture determined by oven drying, often used for milk powder, is known to be affected by the relative air humidity, which has a negative influence on reproducibility. The InfraAlyzer 400 is calibrated with the results of oven drying, in other words, the accuracy of the NIR is highly limited by the accuracy of the reference method. The NIR linear relationship is, however, calculated with a large number of samples spread over a wide range of time. Under these circumstances the influences

TABLE 20.6
NIR Reflectance Data for Protein (g/100 g) in Some Milk Powders

Product	Calibration				Prediction			
	<i>n</i>	<i>R</i>	SEC	Range	<i>n</i>	<i>R</i>	SEP	Range
Skim milk powder	159	0.97	0.27	32.9–41.2	100	0.97	0.25	33.5–40.6
Buttermilk powder	116	0.97	0.21	28.9–34.6	50	0.97	0.21	29.2–34.6
Skim milk powder + buttermilk powder	274	0.98	0.34	28.9–41.2	150	0.97	0.36	29.2–40.6
Denatured milk powder	92	0.98	0.34	27.1–40.1	50	0.96	0.42	28.5–39.7
Milk powder with nonmilk fat	61	1.00	0.18	17.6–26.0	50	0.98	0.25	18.4–25.4

TABLE 20.7
NIR Reflectance Data for Moisture (g/100 g) in Some Milk Powders

Product	Calibration				Prediction			
	<i>n</i>	<i>R</i>	SEC	Range	<i>n</i>	<i>R</i>	SEP	Range
Skim milk powder	159	0.97	0.12	2.9–5.8	100	0.97	0.11	3.1–5.7
Buttermilk powder	116	0.99	0.13	2.7–6.3	50	0.99	0.10	2.7–6.3
Skim milk powder + buttermilk powder	274	0.97	0.13	2.9–6.2	150	0.99	0.14	2.7–6.3
Denatured milk powder	100	0.97	0.11	2.8–6.3	50	0.96	0.16	3.1–6.2
Milk powder with nonmilk fat	64	0.98	0.13	1.3–4.4	50	0.96	0.14	1.5–4.3

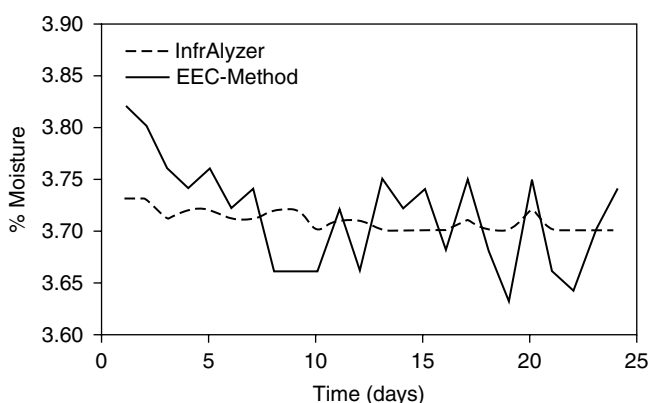


FIGURE 20.5 The relationship of moisture content vs. time for two methods in skim milk powder.

of the relative air humidity are averaged and canceled out. To demonstrate this, 24 tins were filled with a homogeneous sample of spray-dried skim milk powder and air-tight sealed. Moisture in these samples was determined in duplicate on several days over a period of 2 months, according to the oven method and by the InfraAlyzer 400.

Figure 20.5 gives the relationship of moisture content vs. time for the two methods. The reproducibility of the NIR is obviously better than that of the oven method, whereas both methods give the same mean values as expected. Thus the moisture content determined by NIR gives more accurate results than determination by the oven method.

20.4.2.3 Fat

For fat calibrations of skim milk powder it was impossible to get samples with enough variation in fat content. Owing to the standardization of this product the calibration set is not ideal. All calculated correlation coefficients are above 0.98 with an SEC of 0.09 to 0.31% (Table 20.8). The latter 0.31% is found for a combined calibration of skim milk powder and buttermilk powder. The standard error is about twice as high than the separate calibrations, which is caused by an increased variation in the sample composition. On the other hand, samples of milk powder with nonmilk fat give a wide variation of fat content. The SEC value is 0.30. By making two calibrations the SEC value decreases to 0.14 on average. The values of SEP and *R* obtained for the prediction sets are comparable with those of the calibration sets.

The calibration procedure resulted in five to seven filters giving significant information, especially the filters for fat absorption at 2310 and 1759 nm.

20.4.2.4 Lactose

SECs of 0.38, 0.35, and 0.34% were calculated for the determination of lactose in skim milk powder, buttermilk powder, and milk powder with nonmilk fat (Table 20.9). These are acceptable in relation to the standard deviation of the reference method of about 0.25%. The SEC for the combined calibration of skim and buttermilk powder and for denatured milk powder are too high for accurate analysis. Similar results were observed for the test set. This is caused by the variation in physical and chemical composition of the samples and by the few specific NIR bands of lactose in relation to the overall NIR spectrum. Therefore many filters are needed for the calibration. About ten filters gave significant

TABLE 20.8
NIR Reflectance Data for Fat (g/100 g) in Some Milk Powders

Product	Calibration				Prediction			
	<i>n</i>	<i>R</i>	SEC	Range	<i>n</i>	<i>R</i>	SEP	Range
Skim milk powder	159	0.98	0.09	0.8–2.8	100	0.98	0.10	0.5–1.3
Buttermilk powder	116	1.00	0.15	3.3–11.2	50	0.99	0.14	5.3–10.7
Skim milk powder + buttermilk powder	275	0.99	0.31	0.6–11.2	150	0.97	0.41	0.5–10.7
Denatured milk powder	92	0.99	0.19	0.6–7.5	50	0.96	0.24	0.8–6.5
Milk powder with nonmilk fat (overall)	62	1.00	0.30	26.2–50.4	50	0.99	0.34	27.2–49.9
Milk powder with nonmilk fat (lower range)	33	1.00	0.15	32.1–41.5	25	0.99	0.14	32.3–40.9
Milk powder with nonmilk fat (higher range)	29	0.99	0.13	47.4–50.4	25	0.99	0.14	47.9–49.9

TABLE 20.9
NIR Reflectance Data for Lactose (g/100 g) in Some Milk Powders

Product	Calibration				Prediction			
	<i>n</i>	<i>R</i>	SEC	Range	<i>n</i>	<i>R</i>	SEP	Range
Skim milk powder	159	0.95	0.38	42.6–51.6	100	0.94	0.44	43.6–50.2
Buttermilk powder	116	0.99	0.35	37.0–47.6	50	0.99	0.37	37.1–47.3
Skim milk powder + buttermilk powder	275	0.96	0.62	36.6–51.6	150	0.95	0.83	37.1–50.2
Denatured milk powder	92	0.96	0.62	36.8–49.0	50	0.94	0.75	37.5–48.2
Milk powder with nonmilk fat	64	1.00	0.34	23.4–37.6	50	0.98	0.38	24.5–36.1

information. Although each of the filters gave a low correlation coefficient, the combination of ten filters resulted in an acceptable R value.

20.4.2.5 Lactate

Although there are no special claims for the lactate content in buttermilk powder, a calibration for this component was still made (Table 20.10) because it may give information about (a) the original lactose content, (b) the ratio between lactose and protein, or (c) the amount of addition of neutralizing compounds for the lactic acid. The latter influences the ash content.

The lactate content in skim milk powder of about 30 to 100 mg/100 g is too small to be determined by NIR. The combination of skim and buttermilk powder samples in one calibration resulted in an increased accuracy.

For the determination of lactate in buttermilk and denatured milk powder, the use of a lot of filters (12 in total) gives significant information. The need for so many filters is caused by the variation in physical and chemical composition of the samples and by the few specific NIR bands of lactate in relation to the overall NIR spectrum.

20.4.2.6 Ash

In contrast to most organic compounds, most inorganic compounds give no characteristic reflection signals in the NIR region. The correlation between the ash content determined by incineration at 500°C and the ash content predicted by NIRS (Table 20.11) can only be based on the presence of organic and bound milk salts. Probably the ash content predicted by the InfraAlyzer is correlated to the total amount of organic compounds and water because of the large number of filters [7–11] giving significant information.

For skim milk powder the range in ash content is too small for determination by NIRS as can be seen from the relatively high SEC and SEP values and the small correlations coefficients, especially

TABLE 20.10
NIR Reflectance Data for Lactate (g/100 g) in Some Milk Powders

Product	Calibration				Prediction			
	n	R	SEC	Range	n	R	SEP	Range
Buttermilk powder	116	1.00	0.16	0.0–9.0	50	0.99	0.21	0.1–7.2
Skim milk powder + buttermilk powder	275	0.99	0.23	0.0–9.0	150	0.96	0.54	0.0–7.2
Denatured milk powder	92	0.99	0.10	0.1–4.2	50	0.97	0.19	0.3–4.1

TABLE 20.11
NIR Reflectance Data for Ash (g/100 g) in Some Milk Powders

Product	Calibration				Prediction			
	n	R	SEC	Range	n	R	SEP	Range
Skim milk powder	159	0.76	0.08	7.7–8.3	100	0.73	0.11	7.6–8.2
Buttermilk powder	116	0.97	0.19	6.9–10.0	50	0.92	0.23	6.9–9.5
Skim milk powder + buttermilk powder	275	0.85	0.23	6.9–10.0	150	0.85	0.37	6.9–9.5
Milk powder with nonmilk fat	63	0.99	0.09	3.9–6.6	50	0.96	0.15	4.0–6.4
Denatured milk powder	100	0.86	0.18	6.6–9.5	50	0.79	0.27	6.7–9.4

for the prediction set. For the other products quantitative analysis is possible, however, with a relative low accuracy.

20.4.2.7 Skim Milk Powder Content

Skim milk powder can be converted to calf feed by adding nonmilk fat. Because this product is EEC-aided, the determination of the percentage of skim milk powder in calf feed is prescribed by the EEC [29]. Determination of the skim milk powder content is done by enzymatic coagulation of paracasein, a time-consuming method. Moreover, the enzymatic method lacks accuracy. Therefore work has been done to determine the milk powder content by NIRS. The results improve if the skim milk powder content is based on the above-mentioned method together with the results of the protein, lactose, fat, and ash content. For these reasons the calibration of the InfraAlyzer 400 is based on these parameters. The results show that the percentage of skim milk powder in a mixture of milk powder and nonmilk fat can be determined by NIR quantitatively (Table 20.12).

20.4.3 REPEATABILITY AND REPRODUCIBILITY

Repeatability and reproducibility of the determination of moisture, fat, protein, ash, and lactose in skim milk powder with the InfraAlyzer 400 were calculated. For this purpose, 24 subsamples of skim milk powder were analyzed in duplicate on 24 successive working days with the InfraAlyzer 400 (Table 20.13) and with the reference methods (Table 20.14). From Table 20.13 and Table 20.14 it can be concluded that the repeatability of the InfraAlyzer 400 for the determination of moisture,

TABLE 20.12
NIR Reflectance Data for Skim Milk Powder Content (g/100 g) in Milk Powder with Nonmilk Fat

Product	Calibration				Prediction			
	<i>n</i>	<i>R</i>	SEC	Range	<i>n</i>	<i>R</i>	SEP	Range
Milk powder with nonmilk fat	63	1.00	0.43	50.3–76.8	50	0.98	0.55	50.2–75.8

TABLE 20.13
Repeatability and Reproducibility (g/100 g) of InfraAlyzer 400 Method for Skim Milk Powder

	Mean	<i>r</i>	<i>R_w</i>	<i>CV_{R_w}</i>
Moisture	3.710	0.034	0.034	0.3
Fat	0.712	0.037	0.042	2.1
Protein	35.705	0.18	0.18	0.2
Ash	7.964	0.085	0.085	0.4
Lactose	48.779	0.39	0.49	0.4

r = repeatability = $2.83s_r$; R_w = reproducibility over the test period within the laboratory = $2.83s_{R_w}$; CV_{R_w} = coefficient of variation over the test period = $s_{R_w}/\text{mean} \times 100\%$.

fat, and protein in milk powder is the same or better than that of the reference methods, whereas the reproducibility is the same or better for all components determined by NIR.

20.5 CASEIN AND CASEINATES

Caseins are produced by precipitating the protein from milk and drying the precipitate to approximately 8% moisture. Casein is a white crystalline substance and is manufactured in two forms depending on the process used for precipitation: acid casein and rennet casein. Caseins come in various mesh sizes, as well as unground casein. The most common sizes are 30, 60, 80, and 100 mesh. For preparation of sodium or calcium caseinate the casein is mixed with water, sodium hydroxide, or calcium hydroxide and the mixture is spray-dried to approximately 4% moisture. The powder produced is a very fine white powder that is very soluble. Because caseinates have small particle sizes its grinding is not required. Also 60-, 80-, and 100-mesh casein does not require grinding; however, unground casein and 30-mesh casein require grinding for a short time (about 20 s) in a laboratory grinder. The different methods to produce caseine and caseinates not only give products with differences in particle size, shape, and composition, but also with differences in reflectance characteristics (Figure 20.6). Therefore it is better to make different calibrations for each category of product. However, obtaining a relevant number of samples of one category with a range of constituents can be a problem because of the tight specifications demanded. Some analysts therefore manipulate the composition of the calibration samples. A range for moisture can be obtained by drying out some samples and exposing other samples to the atmosphere to absorb moisture. Fat is

TABLE 20.14
Repeatability and Reproducibility (g/100 g)
of Reference Method for Skim Milk Powder

	Mean	<i>r</i>	<i>R_w</i>	<i>CV_{R_w}</i>
Moisture	3.712	0.068	0.15	1.4
Fat	0.718	0.037	0.042	2.1
Protein	35.802	0.207	0.327	0.3
Ash	7.952	0.029	0.075	0.4
Lactose	49.021	0.180	0.478	0.4

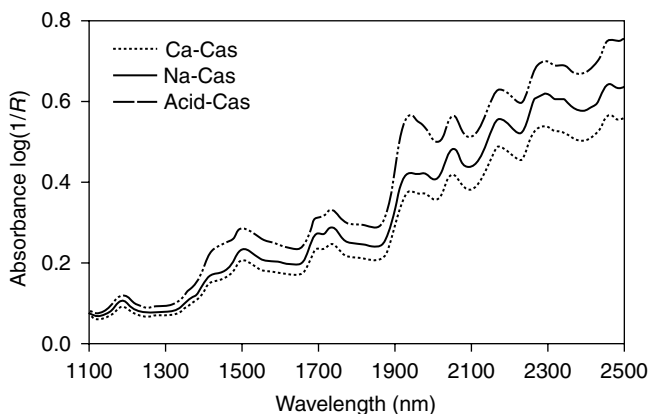


FIGURE 20.6 NIR reflectance spectra of sodium caseinate spray, calcium caseinate spray, and acid casein.

TABLE 20.15
NIR Reflectance Data (g/100 g) Measured with a InfraAlyzer
500 of a Combined Calibration of Casein and Caseinates

Constituent	Calibration				Predication			
	<i>n</i>	<i>R</i>	SEC	Range	<i>n</i>	<i>R</i>	SEP	Range
Moisture	100	0.99	0.14	3.2–8.9	100	0.99	0.15	3.5–8.7
Protein	100	0.95	0.141	85.9–92.9	100	0.93	0.50	86.2–92.8
Fat	100	0.75	0.10	0.52–1.22	100	0.71	0.11	0.6–1.2
Ash	100	0.97	0.25	0.3–4.9	100	0.96	0.27	0.5–4.7

another parameter that remains constant in casein and caseinates. However, manipulating the range by blending high- and low-range samples tends to become unrepresentative of the manufacturing process and therefore will not give accurate results. Another way to obtain samples with a sufficient range for calibration is to manipulate the process of producing casein and caseinates — a rather wasteful and expensive operation.

In order to calibrate our NIRS instruments we have collected during a period of more than a year samples of casein and caseinate originating from the most important casein and caseinate producer in the Netherlands. However, especially for the fat and ash content, it was not possible to get a calibration set with sufficient variation in range that is necessary to make a reliable and robust calibration. In order to get more variation in sample composition we tried to make one “combined” calibration for both casein and caseinates. In spite of the earlier-mentioned differences in particle size and reflectance characteristics between casein and caseinates samples, the results for the calculated parameters were surprisingly good (Table 20.15). Emanating from the required accuracy of 0.1% for moisture, 0.05% for fat, 0.3% for protein, and 0.1% for ash, the accuracy of the determination of the mentioned parameters by NIRS is relative low. However, the NIRS linear relationship is calculated from the results of a large number of samples spread over a wide range of composition. By that, the calibration should be reliable and robust for screening all kinds of casein and caseinates on composition, which is confirmed by testing the calibration with 100 samples of divergent composition (Table 20.15).

20.6 BUTTER

The composition of butter is regulated by law for parameters such as moisture and the solids nonfat content. The economics of manufacture demand that the final products are as near to the target values as possible. Traditionally, therefore, within the industry a wide range of empirical tests have been built up to measure these parameters rapidly in order to be able to correct the production process immediately. Advantage of the NIRS technique is that the very speed of analysis often enables more tests to be carried out, thus providing tighter control than otherwise might be possible. To make a reliable butter calibration for measuring moisture and solids nonfat content, it is important to know that there are a few factors that can influence the accuracy negatively. The first one is the variation in composition. It is not easy to obtain a relevant number of samples with a sufficient range of moisture and solids nonfat content because of the tight specifications demanded. The second factor is the sample temperature. We have observed that a difference in temperature of about 5°C affects the accuracy, negatively caused by differences in absorbance level even after transforming the spectra in second derivative spectra (Figure 20.7a and Figure 20.7b). Therefore, it is absolutely necessary to standardize the sample temperature for NIR analysis. The third important factor is the sample preparation and the loading of the sample cup. Butter that has a very high milk fat content may undergo structural changes by kneading the sample resulting in differences in reflectance. Therefore

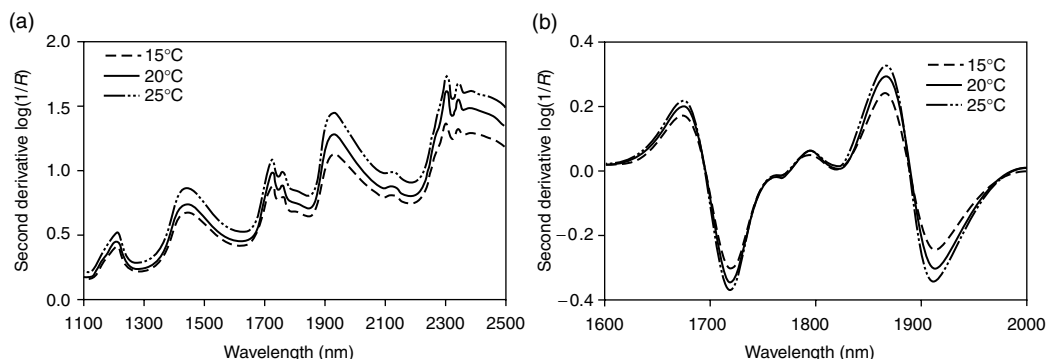


FIGURE 20.7 (a) NIR reflectance spectra for butter obtained by 15, 20, and 25°C. (b) Second derivative NIR-spectra of butter obtained by 15, 20, and 25°C.

TABLE 20.16

NIR Reflectance Data for Moisture and Solids Nonfat Content (g/100 g) of Butter

Constituent	Calibration				Prediction			
	<i>n</i>	<i>R</i>	SEC	Range	<i>n</i>	<i>R</i>	SEP	Range
Moisture	50	0.96	0.07	5.30–16.36	50	0.94	0.08	15.33–16.33
Solids nonfat content	50	0.94	0.07	1.59–2.41	50	0.88	0.09	1.63–2.37

sample preparation and loading of the sample cup must be standardized. Because hardened butter will not spread easily into a cup with a quartz window without kneading of the sample, and because closed cups require more time to clean before measuring of the next sample, we used an open cup. With a little practice it is possible to pack butter samples into an open cup using a spatula and to prepare a surface suitable for determinations of quality or composition.

We have collected 50 samples of hardened butter (sour uncolored) originating from four different factories, and analyzed the samples in duplicate with both the reference methods and NIRs (InfraAlyzer 500). The calibration is tested by selected samples representative of the calibration set as far as possible. The calculated SEC of 0.07% for moisture and solids nonfat content (Table 20.16) are compared with the standard deviation of the reference method of about 0.05%. Concerning the prediction set (test set), the results for the SEP values are comparable with those of the calibration set. It can be concluded that the calibration is useful for screening butter samples on moisture and solids nonfat content. The calibration procedure for moisture resulted in three filters giving significant information, including the filter for water at 1940 nm. The procedure for solids nonfat content selected four filters, including the filter for fat at 1759 nm, for water at 1940 nm, and for protein at 2100 nm.

20.7 CHEESE

Cheese is one of the most difficult products to analyzed using NIRS because of the differences in process, physical and chemical composition, and the high level of moisture and milk fat. Therefore, it is very important to standardize the sampling and sample preparation in order to get an accurate and

representative measurement. The purpose of our study was to investigate the feasibility of applying NIRS to measure some major and minor constituents, as well as the ripening time of the two most important Dutch cheese types: edam and gouda. To this purpose we collected about 200 samples of cheese of both types from six different producers in the Netherlands. The samples were prepared for analysis by removing the rind and ground, using a Hobart mill provided with a cheesehead according to the Dutch standard. The samples were analyzed in duplicate according to the following methods. Moisture content, salt content, and pH were determined according to the Dutch standard. Fat content was measured by the Schmid–Bondzynski–Ratzlaff extraction method. Total nitrogen content was measured according to the Dutch standard making use of a Kjeltec automatic system. The protein content was calculated by multiplying total nitrogen content with the factor 6.38. The soluble N contents were determined according to the method of Noomen [30].

20.7.1 SAMPLE PREPARATION

As a general rule temperature differences of the samples will adversely affect the precision of the results. Products having high milk fat contents may undergo structural changes resulting in differences in reflectance. For similar reasons ground cheese should not be exposed to room temperatures any longer that is strictly necessary and “spreading” movements when filling the sample cup should be avoided.

The RIKILT procedure for preparing cheese samples for NIR is as follows:

1. Take samples according to Dutch standard.
2. Prepare the sample according to Dutch standard. (*Note:* If analysis is postponed, the prepared sample should be stored in a cool place, i.e., about 8°C. It should be removed from cold storage a few hours before commencing the analysis and be allowed to attain room temperature, i.e., $20 \pm 2^\circ\text{C}$, without any active heating. In the sample container the sample should be mixed with a spatula.)
3. Using a spatula, transfer sample to an open sample cup until it is amply filled.
4. Using the spatula, compress the (grained) sample to a level just below the rim of the sample cup. Avoid “spreading” manipulations.
5. Place the sample cup in the sample drawer of the NIR apparatus and close the sample drawer.
6. Carry out the measurement.
7. After measuring, remove the sample material and clean the sample cup.
8. For prediction purposes (composition of unknown samples) make measurements in triplicate (by repeating steps 3 to 6 inclusive).
9. Average the results obtained and calculate by means of the calibration graph the percentages of the components. (*Note:* For drawing the calibration graph it is usually sufficient to perform single measurements.)

20.7.1.1 Effect on Precision

By taking average absorbance/reflectance values of three measurements, a reduction of SEC or SEP of about 30% is obtained. Temperature deviations of 5 to 10°C between the samples produces an increase of the SEC or SEP of 10 to 50%, depending on the component to be determined.

20.7.1.2 Alternative Sample Preparation

In order to reduce the influence of the sample preparation on the accuracy of the NIRS measurements, we investigated an alternative sample preparation (described in the following text). This alternative procedure offers advantages over the previously described one because of the complete elimination

of the variability in grain size, the uniform sample thickness (and thus the uniform height of filling the sample cup), whereas at the same time the (sub)samples will less readily “perspire.” Samples prepared in this manner will produce a reduction in SEC and SEP of 20 to 40% as compared to the SEC and SEP calculated from a similar set of samples prepared by the other method. Nonetheless we have chosen the sample preparation first described for the following two reasons:

1. The sample material is similar to that used for the classical (reference) method.
2. The (sub)samples obtained are considered representative of the entire cheese.

The alternative sample preparation method is as follows:

- Follow the instructions of NEN 3752 up to 4.2.
- With the aid of a slicing apparatus for cheese or meat, cut a 7-mm-thick slice from the sample.
- With the aid of a cork borer, squeeze a small circular piece (of diameter 27 mm) out of the cheese.
- Place this piece in the center of an open sample cup and make the measurement.

20.7.2 RESULTS: MAJOR CONSTITUENTS

As expected, high correlation coefficients between the reference, methods and NIRS were found for moisture, fat, and protein (Table 20.17). The SEC between the reference methods and NIRS are relatively low compared with the required accuracy of 0.2% for moisture, 0.2% for fat, and 0.25% for protein. Better results can be obtained by splitting up the calibration set into two, one for each type of cheese. In that case the SEC values decrease by about 30%. Concerning the results obtained for the prediction sets it can be concluded that the calibrations are robust and reliable for screening samples on major constituents. To give optimal results four wavelengths per constituent were selected.

20.7.3 RESULTS: MINOR CONSTITUENTS

Although NIRS is a new and exciting technique, the main potential of the technique is in the analysis of conventional major constituents such as moisture, fat, protein, and sugars in all kinds of products. Nevertheless we also investigated the possibilities of NIRS to determine some minor constituents in cheese. This was done in order to evaluate the accuracy of NIRS for the analysis of salt and pH for control purposes and to evaluate the possibility of determining parameters such as the amount of nitrogen soluble in water or in trichloroacetic acid (TCA). Preliminary chemical investigations at our institute show that the last-mentioned parameters give a rather high correlation with the ripening

TABLE 20.17
NIR Reflectance Data for Moisture, Fat, and Protein (g/100 g)
of a Combined Calibration of Gouda (48+) and Edam Cheese
(40+) Obtained with an InfraAlyzer 500

Product	Calibration				Prediction			
	<i>n</i>	<i>R</i>	SEC	Range	<i>n</i>	<i>R</i>	SEP	Range
Moisture	125	0.993	0.32	34.4–46.04	100	0.991	0.35	43.11–45.98
Fat	125	0.995	0.26	22.89–43.06	100	0.992	0.28	23.01–34.04
Protein	125	0.995	0.37	23.23–28.45	100	0.995	0.36	23.25–28.40

TABLE 20.18**NIR Reflectance Data for Some Minor Constituents of Gouda (48+) and Edam Cheese (40+)^a**

Constituent	Calibration				Prediction			
	<i>n</i>	<i>R</i>	SEC	Range	<i>n</i>	<i>R</i>	SEP	Range
Salt (NaCl)	125	0.937	0.15	1.2–3.3	100	0.932	0.16	1.2–3.3
pH	125	0.747	0.08	5.1–5.8	100	0.713	0.08	5.1–5.7
Water sol. N/tot. N	125	0.916	2.07	11.1–34.7	100	0.892	2.50	11.1–36.0
TCA sol. N/tot. N	125	0.905	1.34	8.4–22.8	100	0.891	1.51	8.4–24.3
Water-soluble primary amines ($\mu\text{mol/g}$ protein)	70	0.976	43.1	300–160	—	—	—	—

^a Obtained with an InfraAlyzer 500.

age of Dutch cheese [31]. From these it seems possible to control cheese on its declared degree of ripeness in terms of young ripened cheese through measuring the mentioned parameters.

Results for the calculated parameters salt, pH, sol. N/tot. N, TCA sol. N/tot. N, and water-soluble primary amines are surprisingly good (Table 20.18). For salt the second-derivative method gives the best results. This could be expected because the determination of salt by NIRS is based on the change in the behavior of the water component in the spectrum.

Bernal and Fowler [32] showed that the addition of electrolytes changes the spectrum of water in the IR overtone region. Luck [33] showed that temperature variations caused similar spectral changes and attributed these changes to variations in the amount of hydrogen bonding. Begley et al. [34] showed that the best calibration of NIR data to salt content in meat occurs at a point in the spectrum where the salt-induced changes in the water spectrum can be mathematically isolated from other spectral variations by using a scanning NIRS instrument.

The low correlation coefficient of 0.747 for the pH is caused by the small pH range of about 0.6 pH unit. On the other hand, determination of pH in cheese by NIRS is the result of many overlapping overtones that are also very weak. Consequently, it seems impossible to make a high precision calibration for the measurement of pH by NIRS. Mathematical manipulation of the raw NIR data of the cheese samples shows a high correlation or intercorrelation between the parameters — water sol. N/tot. N, TCA sol. N/tot. N, and water-soluble primary amines — and the total protein content. This is to be expected because during cheese ripening proteins are broken down to peptides and amino acids, mainly by enzymatic processes. Because the measurement of protein by NIRS is based on the absorption of casein molecules as well as a variety of peptides and amino acids, it is very difficult to resolve the protein absorption bands in a lot of smaller bands and to correlate these bands to the constituents from which they arise. The only possibility probably is to use higher order mathematical data transformations.

20.7.4 DETERMINATION OF RIPENING TIME

20.7.4.1 General

Dutch regulation for gouda and edam cheese requires a minimum ripening time of 28 days at a minimum temperature of 12°C before they may be sold. This is the only legal constraint set for the degree of ripening of these two types of cheese. For more aged cheese there are no legal rules. Cheese being sold in retail is classified on the basis of its ripening time as follows: young (min. 28 days), young-matured (min. 2 months), matured (min. 4 months), extra matured (min. 7 months), old (min. 10 months), and extra old (min. 1 year). Because ripening conditions for cheese of the same type show little variation, a relationship between maturity (degree of ripening) and ripening time or category can

be expected. Because the cost price is dependent on the category (the younger the cheaper), it is of interest to be able to verify the classification. Preliminary chemical investigations at our institute [31] show that, from an analytical and economical point of view, an acceptable estimation of the ripening time of edam and gouda cheese can be made with a maximum of three independent variables. The three variables chosen by stepwise linear regression analysis were the TCA sol. N/tot. N ratio, which in all cases gave the highest correlation, the salt content, and the moisture content. The proteolytic parameter TCA sol. N/tot. N will increase with time, but the rate of the proteolytic degradation process will diminish with decreasing water content and increasing salt content. Therefore the variance in ripening temperature, humidity, enzyme activity, total composition, etc., influences the relationship between ripening stage and age negatively. The multiple correlation coefficients of the series were situated between 0.94 and 0.95. If discriminating between the young and the young-matured class of both edam and gouda cheese the 95% confidence level is 12 to 13 days, and 18 days if discriminating between the young-matured and matured class.

Next to the above-mentioned chemical investigations, we also studied the suitability of NIRA for the determination of the ripening stage of cheese. The same samples as for the chemical investigations were used. About 100 cheese samples with different ripening stages were collected from different producers. The known age of the cheese samples ranged from 25 to 412 days.

20.7.4.2 Results

For the total sample set we found a correlation coefficient of 0.92 with a SEC of 28 days (Table 20.19). The reason for this high SEC is the very large range of 25 to 412 days for the cheese samples. It seems that in this large range the relationship between ripening stage and age of cheese is not a linear one (Figure 20.8). The SEC is much lower when a calibration with a subset of samples with a range up to 160 days is made. Now a correlation coefficient of 0.96 is found with a SEC of about 11 days, somewhat higher than the value of about 8 days found with the chemical method. Furthermore we have divided the reduced set into two different samples sets because we used two different cheese types: 40% fat in dry matter and 48% fat in dry matter. A multiple correlation coefficient of 0.97 is calculated for both types of cheese with SEC values of 7.2 days (Figure 20.9) and 10 days, respectively (Table 20.19).

It was surprising that the selected wavelengths for the prediction of age are the same as used for predicting the ratio between TCA sol. N/tot. N, the percentage of water, and the percentage of salt. These are exactly the same parameters as are found in the chemical method. Because young-matured cheese is the main product, we tested the combined (40% + 48%) calibration for samples in the range

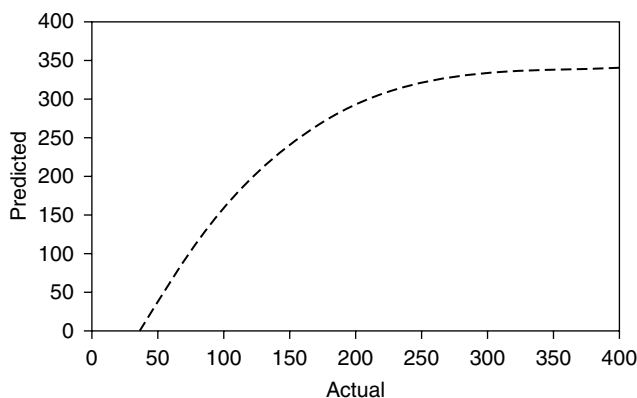


FIGURE 20.8 Relationship between the actual age and the predicted age of two types of Dutch cheese samples (range 25 to 412 days).

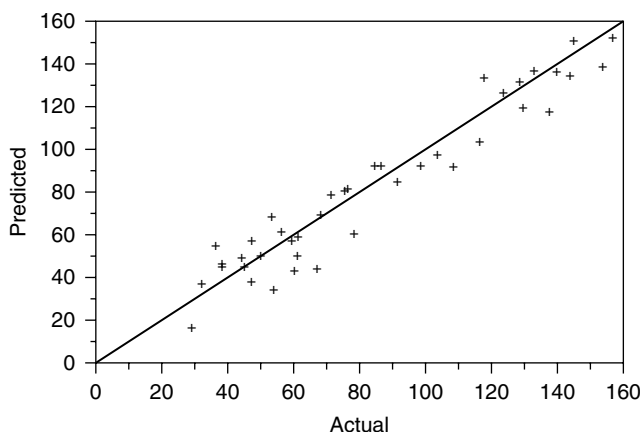


FIGURE 20.9 Relationship between the actual age and the predicted age of cheese samples with 40% fat (range 25 to 160 days).

TABLE 20.19

NIR Reflectance Data for Determination of the Ripening Time (Days) of Edam and Gouda Cheese Obtained with an InfraAlyzer 500

Type of cheese	Calibration			
	<i>n</i>	<i>R</i>	SEC	Range
40% + 48% fat	100	0.92	28	25–412
40% + 48% fat	80	0.96	11	25–160
40% fat ^a	40	0.97	7	25–160
48% fat ^b	40	0.97	10	25–160

^a Gouda cheese with 40% fat in dry matter.

^b Edam cheese with 48% fat in dry matter.

of 25 to 160 days with 20 unknown samples. The results indicated that this calibration is insufficient for quantitative determination of the age of young-matured or matured cheese. Because only the minimum ripening time of a cheese with a significance of 5% ($= 0.05$) is of interest, a one-tail test is required. The 95% reliability interval of the lower tolerance limit, or rejection limit, for the cheese categories is $1.64 \times$ residual standard error. For our test set we calculated for a 95% reliability interval a SEP of about 25 days ($1.64 \times 15 = 25$). As was mentioned previously, the required analytical and economical accuracy in this region must be about 14 days. Therefore we must conclude that determination of the ripening time of cheese with NIRS is insufficient.

20.8 CONCLUSION

The increasing demands of quality control in the dairy industry requires the development of instruments for rapid determination of quality-defining parameters. NIR reflectance instruments have the potential to provide these benefits. The main problem of NIRA is to find a representative sample set to calibrate the NIRA instrument. Sample preparation plays an important role. Particle size, homogeneity, temperature, and presentation of the sample have to be standardized.

The precision of NIRA is limited to a great extent by the precision of the reference methods used for calibration. Therefore, accurate and consistent laboratory analyses are very important.

Useful regression equations can be generated to determine the major constituents in milk, milk powder, casein, butter, and cheese. The accuracy of NIR analysis of major constituents in dairy products is similar to the accuracy obtained with the well-known reference methods. Repeatability and reproducibility of NIRA is the same or better than that of the reference methods.

By using a scanning instrument quantitative determination of parameters such as salt, pH, water sol. N/tot. N, TCA sol. N/tot. N, and water-soluble primary amines is also possible. The accuracy of the determination of these parameters with NIRS is relative low.

Quantitative determination of the age of cheese with NIRS is only possible with low accuracy.

SUGGESTED READINGS

- Ahmed, W. and M. Goldstein, Analytical applications of reflectance spectroscopy in the near infrared region: A method for quality control of fat in dried milk, *Lab. Pract.*, 25: 385 (1976).
- Anon., Applications note on measurement of fat and total solids in milk samples using a GQA-31 EL, Pacific Scientific Technical Paper NIR 4000, May 18, 1978.
- Anon., Infrared milk analysis, Interim official first action, Official Methods of Analysis of the Association of Official Analytical Chemists, Ed. Horwitz, W. AOAC, Washington, DC, 13th ed., 1980, pp. 248–249.
- Atherton, H. V. and J. A. Newlander, Chemistry and testing of dairy products, AVI Publ. Co. Inc. Westport, CT, 4th ed., 1980, pp. 105–111.
- Biggs, D. A., Instrumental infrared estimation of fat, protein and lactose in milk: Collaborative study, *J. Assoc. Off. Anal. Chem.*, 61: 1015 (1978).
- Biggs, D. A., Performance specifications for infrared milk analysis, *J. Assoc. Off. Anal. Chem.*, 62: 1211 (1979).
- Biggs, D. A., Infrared estimation of fat, protein and lactose in milk: Evaluation of Multispec instrument, *J. Assoc. Off. Anal. Chem.*, 62: 1202 (1979).
- Black, R. G., Analysis of raw milk by NIR, *Proc. NIR 84: An International Symposium on Near Infrared Reflectance Spectroscopy*, Melbourne, Australia, 1984, pp. 105–114.
- Blanco, C., P. Casado, and A. Poza, Application of diffuse reflectance spectroscopy. It's application to the determination of powder milk composition, *Quim. Ind. (Madrid)*, 26: 361–363 (1980) (in Spanish).
- Brown, R. J., Modern methods of analysis of milk and milk products: Rapid determination of main components, milk: The vital force, 507 (1987).
- Collyer, M. J., An evaluation of the Neotec model 102 microcomputer controlled scanning spectrophotometer, for the measurement of major component in full cream milk powder. Dissertation submitted to the University of Reading, Department of Food Science, in partial fulfillment of the degree of Master in Science in Dairy Science, September 1982.
- Diller-Zulauf, A. and H. R. Egli, Quantitative rapid determination of fat, nitrogen and dried substances in whey products with the InfraAlyzer, *Dtsch. Molk.-Ztg.*, 103: 820–822 (1982) (in German).
- Diller-Zulauf, A. and H. R. Egli, Quantitative Schnellbestimmung von Lactose, Fett-Stickstoff- und Trockensubstanz an Milch sowie des MilCHFettgehaltes von Rahm mit dem InfraAlyzer 400 Dairy (DR), *Alimenta*, 5: 107; 6: 148 (1985).
- De Vilder, J. and R. Bassuyt, Practical experiences with an InfraAlyzer 400 in determining the water, protein and fat content of milk powder, *Milchwissenschaft*, 38: 65 (1983).
- Egli, H. R., Standardisation, quality insurance and control of milk products with the Technicon InfraAlyzer 400 Dairy, *NIRA International Symposium*, Technicon, Paris, November 1983.
- Egli, H. R. and U. Meyhack, Measurements of the Principal Constituents of Solid and Liquid Milk Products by Means of Near Infrared Analysis, *Proceedings of a Seminar*, University of Reading, England, March 28–30, 1984. London, Royal Society of Chemistry, 1984.
- Frank, J. F., Möglichkeiten zur Schnellbestimmung der Käsezusammensetzung durch die Anwendung der Reflexionsspektroskopie, *DMZ: Dtsch. Molk.-Ztg.*, 15: 438 (1983).
- Gilkison, I., An evaluation of the Technicon InfraAlyzer 400 Dairy for the analysis of milk and dairy products, *NIRA International Symposium*, Technicon, Paris, November 1983.
- Goulden, J. D. S. and D. J. Manning, Determination of moisture in dairy products by near-infrared absorption of methanol extracts, *J. Dairy Res.*, 37: 107 (1970).

- Guiader, M., Cheese and meat analysis by NIR, *International Near Infrared Diffuse Reflectance/Transmittance Spectroscopy Conference*, Budapest, Hungary, May 1986.
- Hammingh, A., Experience with the Technicon InfraAnalyzer 400 R in dried dairy products, as well as on-line analysis of milkpowders, *NIRA International Symposium*, Technicon, Paris, November 1983.
- Haugaard, C., Photometric determination of fat in milk, *J. Dairy Sci.*, 49: 1185 (1966).
- Hirschfeld, T., Sample area optimization in a diffuse reflectance near-infrared spectrophotometer, *Appl. Spectrosc.*, 39: (1985).
- Honigs, D. E., Near infrared analysis, *Anal. Instr.*, 14: 1–62 (1985).
- Honigs, D. E., G. M. Hieftje, and T. Hirschfeld, Number of samples and wavelengths required for the training set in near-infrared reflectance spectroscopy, *Appl. Spectrosc.*, 38(6): (1984).
- Kaffka, K. J. and F. Kulscár, Attempts to determine egg content in pastry products using the NIR-technique, *Acta Alimentaria*, 11: 47 (1982).
- Lawrence, R. C. and J. Gilles, The assessment of the potential quality of young Cheddar cheese, *N. Zeal. J. Dairy Sci. Technol.*, 15: (1980).
- Luchter, K., Interlaboratory milk study with the Technicon InfraAnalyzer 400 Dairy, *NIRA International Symposium*, Technicon, Paris, November 1983.
- McGann, T. C. A., Control of moisture in skim milk powders by IR reflectance using the Anocan model 106, *Anim. Prod.*, The Agriculture Institute, 1975.
- Mills, B. L. and F. R. Van de Voort, Evaluation of CH stretch measurement for estimation of fat in aqueous fat emulsions using infrared spectroscopy, *J. Assoc. Off. Anal. Chem.*, 65: 1357 (1982).
- Mitchell, D. J., R. W. Weik, and W. Hortritz, Collaborative studies of methods for butterfat in homogenized and chocolate milk, *J. Assoc. Off. Agric. Chem.*, 47: 573 (1964).
- Roy, R. B. and J. Bahl, Theoretical Basis for the Analysis of Fats and Oils by NIRA, *Presented at the 2nd Annual NIRA Symposium*, Technicon Corporation, Tarrytown, New York, June 1982.
- Rudzik, L. and R. Wobbecke, Need for homogenization in infrared measurements on milk, *DZM: Dtsch. MolK.-Ztg.*, 36: 298 (1982) (in German).
- Sjaunja, L.-O., Studies on milk analysis of individual cow milk samples, thesis, Swedish University of Agriculture Sciences, 1982.
- Toohy, F. S., Testing of dried milk by NIR, *Proc. NIR 84: An International Symposium on Near-Infrared Reflectance Spectroscopy*, Melbourne, Australia, 1984, pp. 114–119.
- Whetsel, K. B., Near-infrared spectrophotometry, *Appl. Spectrosc. Rev.*, 2: 1–67 (1968).
- Wollard, D. C., NIR technologie within the New Zealand dairy Industry: Its dramatic impact, *Proc. NIR 84: An International Symposium on Near-Infrared Reflectance Spectroscopy*, Melbourne, Australia, 1984, pp. 95–105.

REFERENCES

1. G. M. Roberts, Infrared Analysis of Milk Powders and Casein Products. Technical paper Pacific Scientific No. 4007, 1977.
2. R. W. V. Weaver, Infrared Reflectance Analysis Applied to the Dairy Industry, *Proc. 8th Technicon Intl NIRA Congress*, London, 1978.
3. R. W. V. Weaver, Near Infrared Reflectance Analysis Applied to Dairy Products, *Proc. of a Seminar*, University of Reading, London. Royal Society of Chemistry, 1984, pp. 91–102.
4. A. M. C. Davies and A. Grant, *Int. J. Food Sci. Technol.*, 22: 191 (1987).
5. T. C. A. McGann, *Irish J. Food Sci. Technol.*, 2: 141 (1978).
6. D. A. Biggs, *J. Assoc. Off. Anal. Chem.*, 55: 488 (1972).
7. D. A. Biggs, Applications of Infrared Instrumentation in Cheese Analysis, *Proc. 1st Biennial Marshall Int. Cheese Conference*, Madison, Wisconsin, 1979, pp. 409–414.
8. J. D. S. Goulden, *J. Dairy Sci.*, 24: 242 (1957).
9. R. J. Bear, J. F. Frank, and M. Loewenstein, *J. Assoc. Off. Anal. Chem.*, 4: 66 (1983).
10. I. Ben-Gera and K. H. Norris, *Isr. J. Agric. Res.*, 18: 117 (1968).
11. P. Casado, C. Blanco, A. Pozas, and L. Matorras, *Proc. Intl Dairy Congress*, 1978, p. 376.
12. R. J. Bear and J. F. Frank, Analysis of Nonfat Dry Milk Using Near Infrared Reflectance Spectroscopy, *Proc. 2nd Annual NIRA Symposium*, Technicon, Tarry-town, New York, 1982.

13. C. H. White, G. A. Muck, M. Bulthaus, and P. Rotolo, The Use of IR Reflectance Analyzer for Measurement of Fat and Total Solids of Dairy Products, *Proc. 73rd Annual Meeting of Am. Dairy Science Assoc.*, East Lansing, MI, 1978.
14. G. S. Birth and C. J. Washam, *J. Dairy Sci.*, 60: 57 (1977).
15. R. Giangiacomo, D. Torreggiani, J. F. Frank et al., *J. Dairy Sci.*, 62: 39 (1979).
16. F. J. Frank and G. S. Birth, *J. Dairy Sci.*, 65: 1110 (1982).
17. R. Frankhuizen, E. A. M. Boers, and H. Oortwijn, *Zuivelzicht*, 75: 210 (1983).
18. R. Frankhuizen, E. A. M. Boers, and H. Oortwijn, *Zuivelzicht*, 75: 547 (1983).
19. R. Frankhuizen, The Use of NIRA for Quality Control of Dairy Products, *Proc. 8th Int. Symposium on Near Infrared Reflectance Analysis*, Technicon, Tarrytown, New York, 1985.
20. R. Frankhuizen and N. G. van der Veen, *Neth. Milk Dairy J.*, 39: 191 (1985).
21. R. Frankhuizen, The Use of NIRS for Quality Control of Dairy Products, *Proc. Intl Near Infrared Diffuse Reflectance/Transmittance Spectroscopy Conference*, Budapest, 1986.
22. M. Tuinstra-Lauwaars, E. Hopkin, and S. Boelsma, *J. Assoc. Off. Anal. Chem.*, 68: 1235 (1985).
23. J. D. S. Goulden, *J. Dairy Res.*, 31: 273 (1964).
24. H. Meistre, *Proc. Cold Spring Harbor Symp. Quant. Biol.*, Vol. 3, 1935, pp. 191–209.
25. W. L. Butler and K. H. Norris, *Arch. Biochem. Biophys.*, 87: 31 (1960).
26. P. Latimer, *Plant Physiol.*, 32: 17 (1957).
27. K. H. Norris, Measuring and Using Transmittance Properties of Plant Materials, *Proc. Electromagnetic Radiation in Agriculture Conference*, Illuminating Engineering Society, New York, and American Society of Agricultural Engineers, St. Joseph, Michigan, 1965, pp. 64–66.
28. D. Bertrand, P. Robert, D. Launay, and M. F. Devaux, Application of Principal Component Analysis to the Determination of Fat and Protein Content of Milk by Near Infrared Spectroscopy, *Proc. Euro Food Chem. IV*, Leon, Norway, 1987, pp. 519–523.
29. EEC regulation nr. 1725/79. Annex III, method L296/10. Quantitative analysis of skimmed milk powder in feedingstuffs by enzymatic coagulation of casein.
30. A. Noomen, *Neth. Milk Dairy J.*, 31: 163 (1977).
31. D. P. Venema, H. Herstel, and H. L. Elenbaas, *Neth. Milk Dairy J.*, 41: 215 (1987).
32. J. D. Bernal and J. Fowler, *Chem. Phys.*, 1: 515 (1933).
33. W. A. P. Luck, *Structure of Water and Aqueous Solutions*, Hans Richaz, St. Augustine, West Germany, Ch. 3, 1974.
34. T. H. Begley, E. Lanza, K. H. Norris, and W. R. Hruschka, *J. Agric. Food Chem.*, 32: 984 (1984).

21 Near-Infrared Spectra of Gases

Chris W. Brown

CONTENTS

21.1	Introduction	439
21.2	Comparison of MIR and NIR Spectra of Gases	440
21.3	Substitution Patterns in NIR Spectra.....	441
21.4	Comparison of NIR Spectra of Gases and Solutions.....	441
21.5	Monitoring Natural Gas Compositions and BTU Contents.....	443
21.5.1	NIR of Natural Gas	444
21.5.2	Resolution Required for Quantitative Predictions	446
21.5.3	NIR Field Measurements of Natural Gas.....	447
21.6	NIR Gas-Phase Library and Search with Mixtures	448
21.6.1	Preparation of the Spectral Library	448
21.6.2	Search Algorithm	449
21.6.3	Search Results	450
21.6.4	Summary of the Search Results.....	454
21.7	Conclusions	455
	References	455

21.1 INTRODUCTION

Near-infrared (NIR) spectroscopy has been used extensively in agriculture, pharmaceutical, and process control areas [1–8]. It is an ideal method for determining moisture and protein content of food products [1–3], composition and thickness of polymer films [5], and octane ratings of gasoline [6,8]. More and more practical uses of NIR are discovered each year. However, NIR applications to gases have been very limited. Mid-infrared (MIR) has been used to monitor the composition of environmental gases for many years [9,10]. The major difference between MIR and NIR of gases is in the sensitivity. Absorptivities in the MIR are 10- to 100-fold stronger than those in the NIR. There has also been some concern about the selectivity of NIR vs. MIR, since bands tend to be broader and there is more overlap in the NIR region.

Like other practical applications of NIR, eventually this technology will make a major impact on gas monitoring. There have already been several attempts to monitor environmentally significant gases using diode lasers emitting at ~ 1.3 or ~ 1.5 μm , which are the types of lasers used extensively for fiber optic communication systems. Applications to gaseous molecules are confined to those molecules having absorptions in the same regions as the diode lasers [11]. The investigations have included those on water vapor [12], carbon monoxide [13], carbon dioxide [11,13,14], and methane [11,15–17]. Most of these studies have been devoted to determining the feasibility of diode lasers to detect low concentrations of the analyte.

In this chapter, we will consider spectral contours of gases in the NIR and compare them with those in the MIR and with those in solutions. It will be shown that the NIR fingerprint region is very similar to that in the MIR. These comparisons will be followed by a quantitative application of NIR for determining the compositions and energy contents of natural gas. Finally, it will be shown that NIR can be used effectively in searching spectral libraries with unknown spectra of mixtures.

21.2 COMPARISON OF MIR AND NIR SPECTRA OF GASES

MIR and NIR spectra of five typical organic gases are compared in Figure 21.1 and the respective pressures are compared in Table 21.1. The MIR spectra were measured in a cell having a pathlength of 1.8 m, whereas the NIR spectra were measured in a 20 m cell. Other than the need for higher pressure and longer pathlengths, the spectra are similar for the two spectral regions. Characteristic

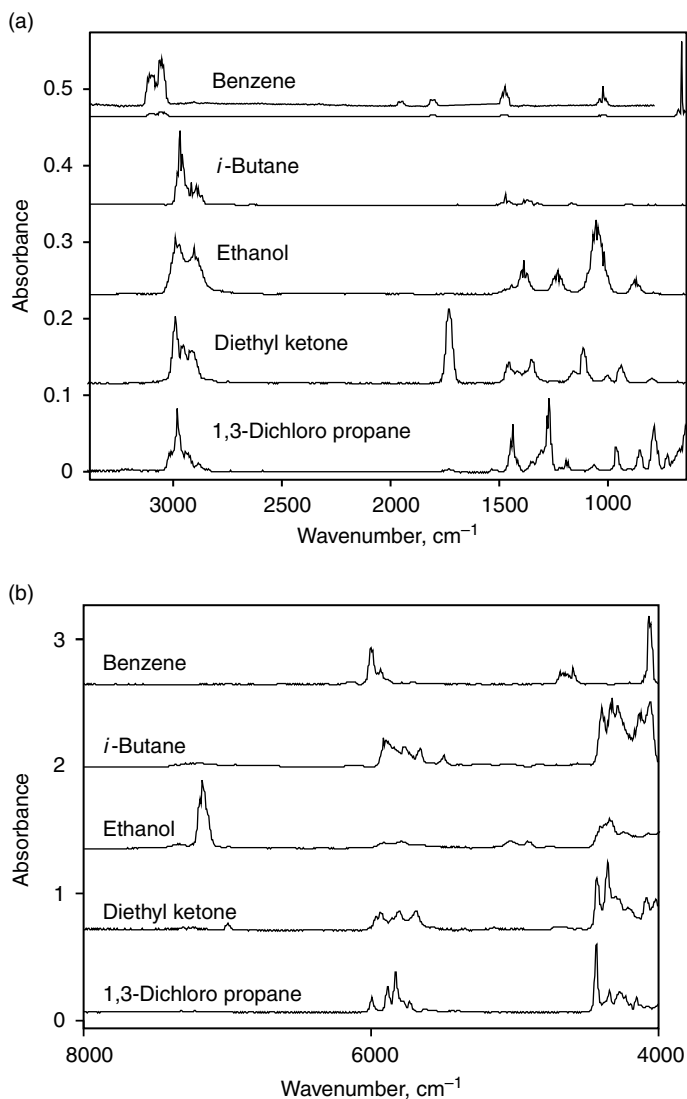


FIGURE 21.1 (a) MIR spectra of typical environmentally important gases. (b) NIR spectra of the same gases.

TABLE 21.1
Comparison of Pressures Used for Measuring MIR and NIR Spectra of Organic Gases

Organic compound	Pressure (torr) for MIR	Pressure (torr) for NIR
	1.0 m pathlength	20 m pathlength
Benzene	10	42
<i>i</i> -Butane	148	152
Ethanol	320	65
Diethyl ketone	320	36
1,3-Dichloro propane	300	40

absorptions due to C—H stretches appear at $\sim 3000\text{ cm}^{-1}$ in the MIR and at $\sim 6000\text{ cm}^{-1}$ in the NIR. Even more important is that spectra in the NIR have a fingerprint region from 4000 to 5000 cm^{-1} , which is similar to the MIR fingerprint region of 600 to 1800 cm^{-1} . Absorptions in the 4000 to 5000 cm^{-1} are due to combinations of fundamental vibrations in the MIR fingerprint region with fundamental C—H stretching vibrations.

For combination and overtone bands to be active, a particular vibration must be anharmonic, since combinations and overtones are not allowed in the harmonic oscillator approximation. (In the harmonic oscillator, the vibrational energies are related to the square of the bond displacements during the vibration, whereas in the anharmonic oscillator, the same energies are related to higher powers of the displacements.) Vibrations exhibiting large anharmonicities can have reasonably strong bands in the NIR region. Thus, relative intensities of bands observed in the MIR may not carry over into the NIR, that is, the relative intensities for functional groups may very well change. For example, a certain molecule may have a stronger spectrum than another molecule in the MIR, but the spectral strengths may reverse in the NIR.

21.3 SUBSTITUTION PATTERNS IN NIR SPECTRA

The NIR fingerprint region from 4000 to 5000 cm^{-1} exhibits distinctive features, which can be used for identifying a particular unknown; however, there are also some characteristic contours for similar substitution patterns. This effect is shown in Figure 21.2 for alkyl-substituted aromatics. All aromatics exhibit a strong doublet centered at 4050 to 4100 cm^{-1} and a weaker doublet (or triplet) centered at $\sim 4650\text{ cm}^{-1}$.

Spectra of six chlorinated alkanes are shown in Figure 21.3. It is interesting that although the C—Cl stretching fundamnet is very low (at $\sim 700\text{ cm}^{-1}$) in the MIR, the chlorinated substitution pattern forms a very distinctive fingerprint in the NIR. The di-substituted ethane, propane, and butane have similar patterns in the 4300 to 4500 cm^{-1} region, but there are enough other differences to clearly identify the different molecules.

21.4 COMPARISON OF NIR SPECTRA OF GASES AND SOLUTIONS

NIR spectra of five gases and solutions of the same organic compounds are compared in Figure 21.4. The gas-phase spectra were measured in a 20 m cell. The solution spectra were measured using carbon tetrachloride as the solvent and the concentrations of the solute were adjusted to produce maximum absorptions of ~ 1.0 . There are two noticeable differences between the gas and solution spectra; the relative band intensities are somewhat different, and some of the bands in the gas-phase spectra are sharper than those in the solution. The intensities of the bands in the C—H overtone

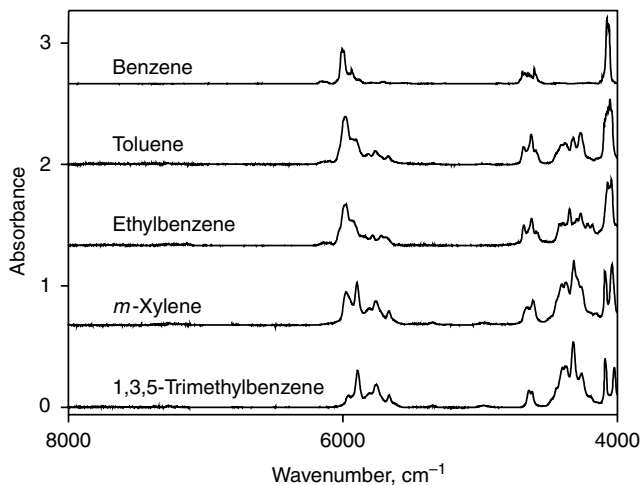


FIGURE 21.2 NIR spectra of substituted aromatic gases.

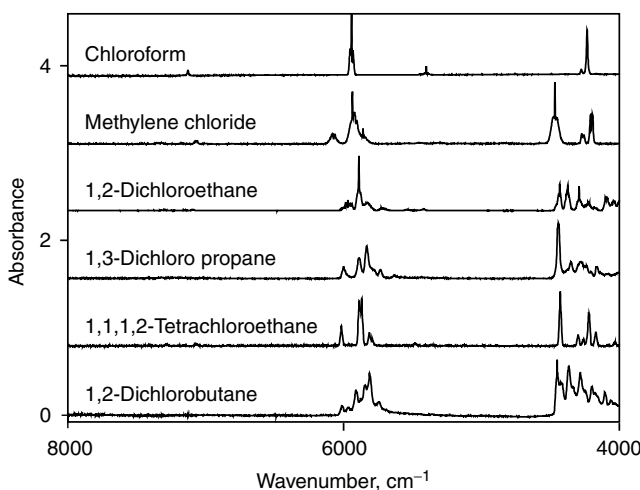


FIGURE 21.3 NIR spectra of chlorinated alkanes.

stretching region ($\sim 6000\text{ cm}^{-1}$) relative to those in the $4000\text{ to }5000\text{ cm}^{-1}$ fingerprint region are stronger in the gas phase for all of the compounds except hexane. The number of bands and their positions change very little in going from solutions to gases.

We might expect to see the greatest difference between gas phase and solution spectra for small molecules and for molecules that are susceptible to association such as hydrogen bonding. Both of these effects are apparent in the spectra of methanol shown in Figure 21.5. The gas-phase spectrum exhibits a doublet in the —OH stretching overtone region at $\sim 7200\text{ cm}^{-1}$. This doublet is due to the superposition of rotational fine structures corresponding to increasing and decreasing rotational quantum numbers for the gaseous molecules. In the solution phase, the sharp band at 7022 cm^{-1} is due to the first overtone of the OH stretch of a nonhydrogen bonded monomer. The very broad band extending from $7000\text{ to } \sim 6100\text{ cm}^{-1}$ is due to hydrogen-bonded oligomers. The combination band at 5025 cm^{-1} in the gas-phase spectrum shifts to $\sim 4800\text{ cm}^{-1}$ and becomes noticeably broader in the solution spectrum. Finally, there is a large increase in the band intensity of the band at 4390 cm^{-1} . Thus, both the size of the molecule and molecular interactions greatly affect the spectral contours.

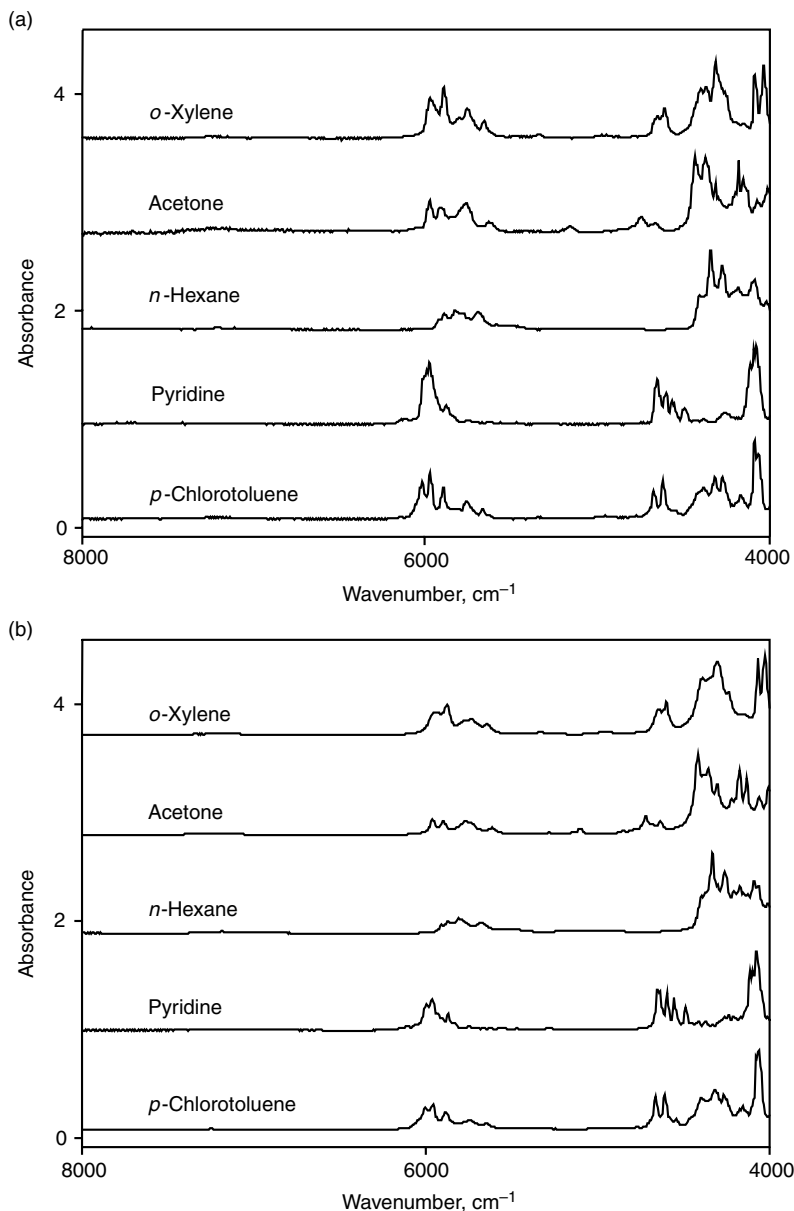


FIGURE 21.4 (a) NIR spectra of typical gases. (b) NIR spectra of solution of the same compounds in CCl₄ using CCl₄ for the background spectrum.

21.5 MONITORING NATURAL GAS COMPOSITIONS AND BTU CONTENTS

One of the few practical applications of NIR to gases has been its proposed use as a monitoring system for natural gas. In the United States, the composition of natural gas and its British Thermal Unit (BTU) content in transmission lines is measured at each state boundary. At present, gas samples are taken off-line and analyzed automatically by gas chromatography (GC). In the case of several transmission lines crossing at a specific location, a system is set up to automatically sample each of the lines periodically. Flow meters are then used to measure the volume of gas consumed by each customer.

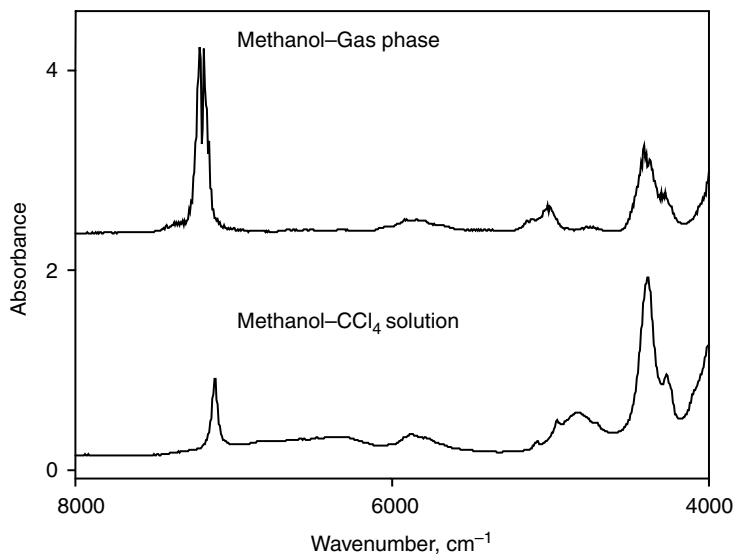


FIGURE 21.5 NIR spectra of methanol in the gas phase and in CCl₄.

TABLE 21.2
Range of Concentrations for
Components of Natural Gas

Component	Range (%)
Methane	85 \Rightarrow 97
Ethane	0.5 \Rightarrow 15.0
Propane	0.2 \Rightarrow 3.0
<i>n</i> -Butane	0 \Rightarrow 0.42
<i>i</i> -Butane	0 \Rightarrow 0.33
<i>n</i> -Pentane	0 \Rightarrow 0.12
<i>i</i> -Pentane	0 \Rightarrow 0.11
Hexanes	0 \Rightarrow 0.16
CO ₂	0 \Rightarrow 2.1
N ₂	0 \Rightarrow 4.0

The volume of gas consumed and its pressure are used along with the composition to determine the BTUs, and the customer pays for the BTUs. This might seem like a straight forward process and it would be so if the composition measured at each state boundary remained constant over time.

The ranges of concentrations for components of natural gas are listed in Table 21.2. The major component is methane and its concentration can range from a low 85 to as high as 97%. Spectra of natural gas are dominated by methane. However, the larger alkanes contribute more energy per molecule to the BTU content. Each CH₃— (or —CH₂—) group in a molecule contributes approximately the same energy in terms of BTUs.

21.5.1 NIR OF NATURAL GAS

NIR spectroscopy is ideal for monitoring the composition and BTU content of natural gas. Pressure in gas transmission lines is on the order of 400 psi. At this pressure, absorptions in the MIR are too high at reasonable pathlengths. In the NIR, natural gas has absorptions at 1100–1200, 1325–1425,

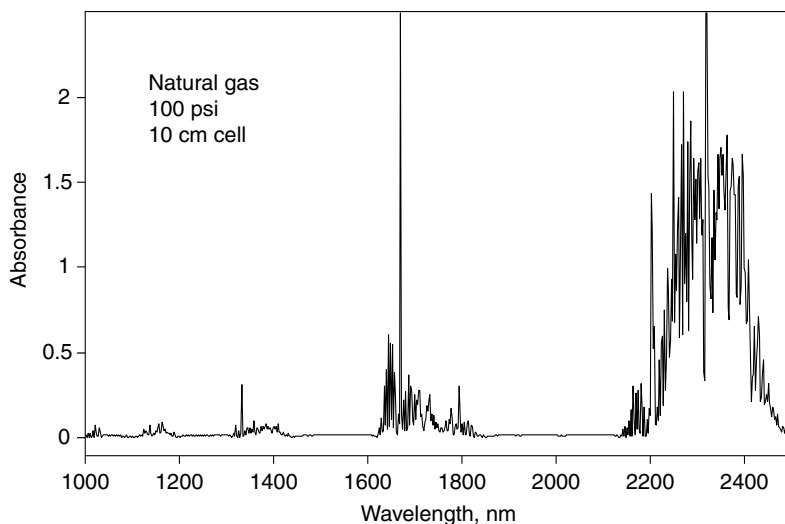


FIGURE 21.6 NIR spectrum of simulated natural gas at 100 psi in a 10-cm cell.

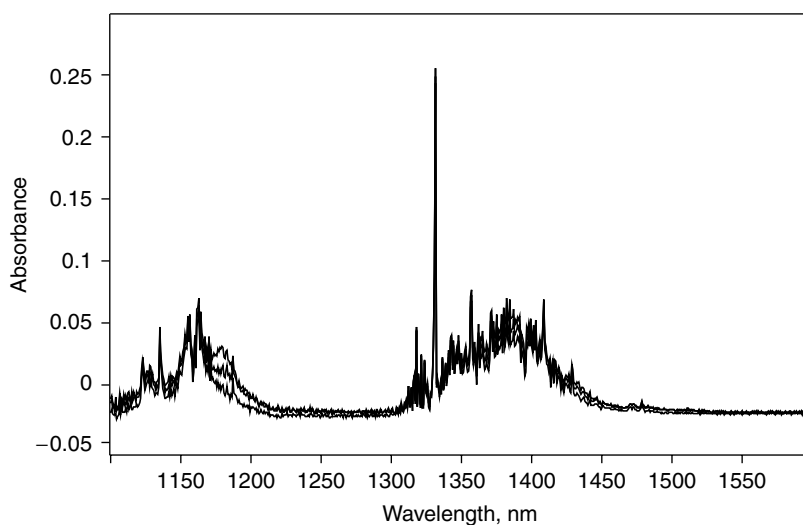


FIGURE 21.7 NIR spectra of simulated natural gas samples containing the same amount of methane with varying amounts of ethane and propane.

1600–1800, and 2200–2500 nm as shown in Figure 21.6: this spectrum was measured at 100 psi in a 10-cm pathlength cell. The bands at 1100–1200 and 1325–1425 nm have appropriate absorptions for the 100 to 600 psi range at a pathlength of 10 cm.

NIR spectra from 1100 to 1600 nm of natural gas having three different compositions are shown in Figure 21.7. These samples all contain ~85% methane, 2–15% ethane, and 0–5% propane. Variations caused by the different amounts of ethane and propane are observed at ~1175 and ~1440 nm; however, the dominance of methane is obvious. The difficulty in using NIR spectra to monitor the composition and energy content of natural gas is that methane is the major contributor to the spectra, but the other alkanes provide greater energy. The spectrum of methane is composed of many sharp vibrational–rotational lines as shown in Figure 21.7. The spectral contributions of ethane, propane, and the higher alkanes are much broader and less pronounced.

21.5.2 RESOLUTION REQUIRED FOR QUANTITATIVE PREDICTIONS

In an attempt to design an inexpensive system for monitoring natural gas in the early 1990s, we explored the feasibility of using a lower resolution instrument. The design of a practical field system for monitoring natural gas was limited by the availability of array detectors at that time and cost considerations. As discussed above, the optimum spectral region was 1100 to 1600 nm for gas pressures of 100 to 600 psi and a gas cell pathlength of 10 cm. There was also a requirement that the equipment should not have any moving parts other than solenoid-operated gas valves. This requirement mandated a nonscanning spectrometer fitted with a linear array detector. Germanium arrays were the best choice at that time, but how many elements would be required in the array, that is, what was the required resolution? In addition to these requirements, the instrument had to compete with gas chromatographic methods, which had reported precisions and accuracies of $<0.1\%$. Thus, the problem was to develop an NIR monitoring system, which could compete with GC instrumentation and produce similar results in real times for less money.

To determine the effect of resolution and the required number of linear array detector elements, we measured spectra of 24 simulated natural gas samples. These samples were prepared by weight at the University of Oklahoma [18]. Spectra of these samples were measured in the NIR using a FTIR at a resolution of 2 cm^{-1} (or $\sim 4\text{ nm}$). The original spectra were analyzed by principal component regression (PCR) using cross-validation to develop a prediction model. The model was used to predict the BTU content of each sample and the concentrations of each of the nine components. These original spectra were then convoluted to 10, 20, \dots , 100 nm using a Gaussian-shaped slit function. Each of these sets of convoluted spectra was analyzed by PCR; prediction errors as a function of convolution are shown for the BTU and some of the major chemical components for selected sets in Figure 21.8. The minimum errors on the BTU content and concentrations of methane and pentanes were obtained at a resolution of $\sim 20\text{ nm}$.

In designing a nonscanning spectroscopic monitoring system, two effects have to be considered. In addition to resolution, the number of detector elements also influenced the spectra and the predictions. The effect of spectral resolution on the spectra is shown in Figure 21.9, whereas the

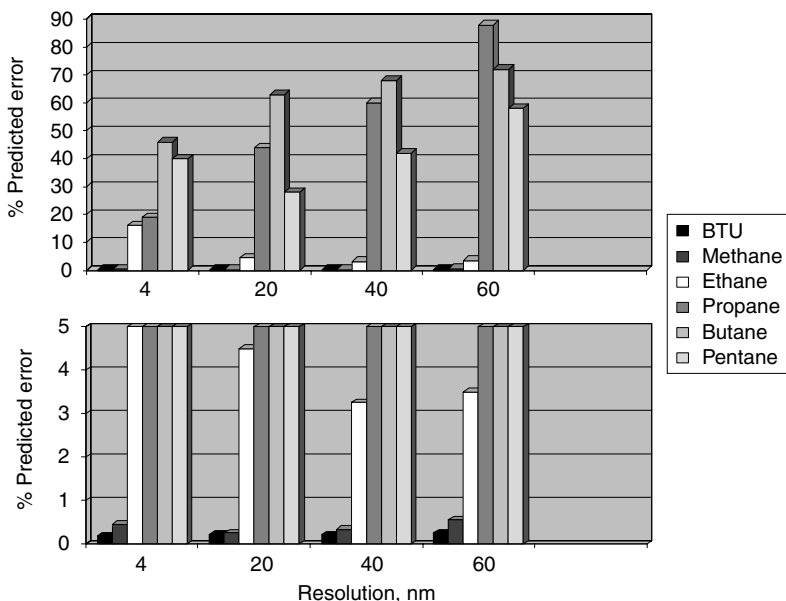


FIGURE 21.8 Error bars for the cross-validation prediction of 24 standard simulated natural gas samples as a function of resolution.

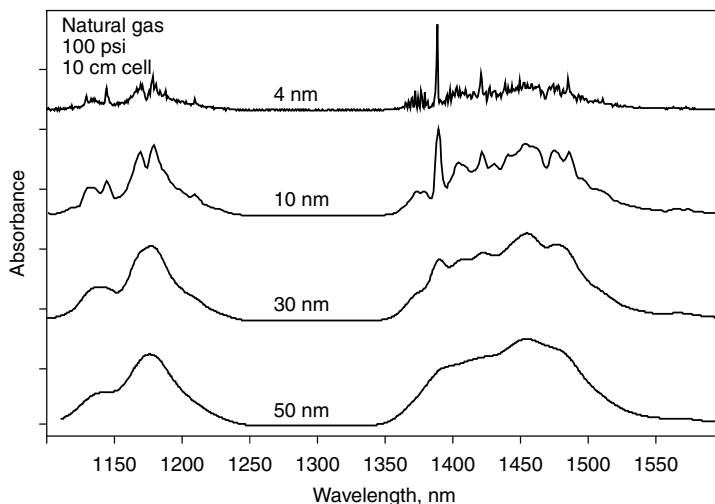


FIGURE 21.9 Simulated natural gas spectra as a function of resolution.

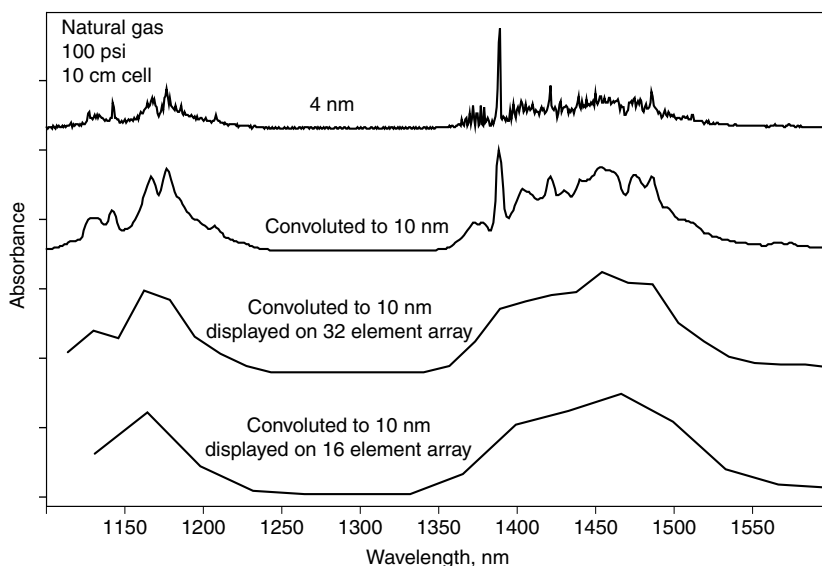


FIGURE 21.10 NIR spectrum of a simulated natural gas sample at 4 nm resolution, convolved to 10 nm, and displayed on a 32-element array and on a 16-element array detector.

effect of the number of array elements is shown in Figure 21.10. For the spectral range of 1100 to 1600 nm, we found that a theoretical spectral resolution of 16 nm displayed on a 32-element array produced optimum results. In the final system, a 32-element Ge-diode array was used with an American Holographic Monochromator as the detection system [19]. The spectra obtained with this prototype instrument were very similar to the third spectrum in Figure 21.10, that is, to the 4 nm spectrum convolved to 10 nm and displayed on a 32-element linear array.

21.5.3 NIR FIELD MEASUREMENTS OF NATURAL GAS

To monitor natural gas using NIR, an off-line system was constructed as is diagramed in Figure 21.11 [20]. Gas from a main line was bled into a high-pressure cell. For preliminary measurements, the gas

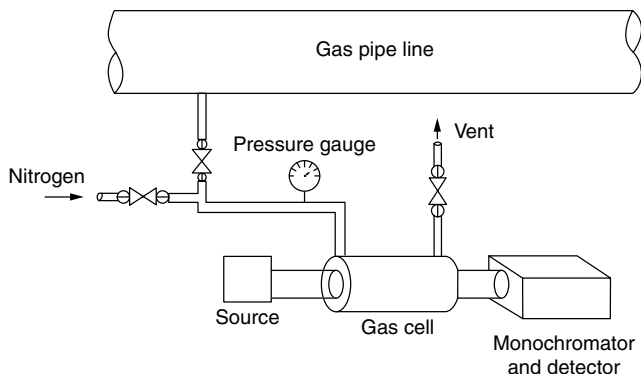


FIGURE 21.11 Schematic of the off-line sampling system for monitoring natural gas transmission lines.

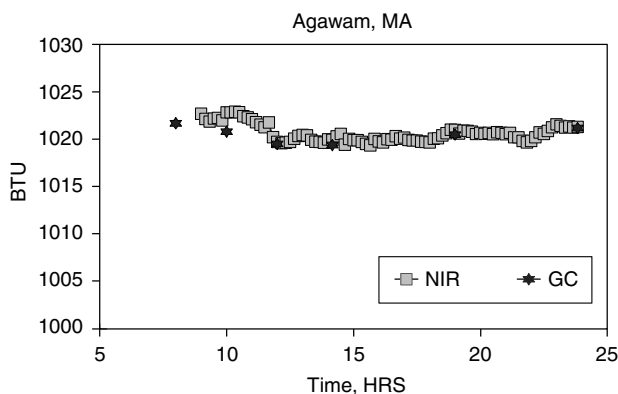


FIGURE 21.12 BTU of natural gas predicted by off-line analysis using the NIR monitor vs. gas chromatography.

pressure was reduced to 35 psi in a 10-cm pathlength cell. The spectral resolution was ~ 16 nm for the range of 1100 to 1600 nm. The monitoring system was calibrated with the same 24 simulated gas samples containing nine chemical components as discussed above. PCR was used to model the data and to predict the component concentrations and the BTU contents. The same off-line gas samples were also analyzed by gas chromatography every 2 h. The NIR measurements were made once each minute, the spectrum of natural gas was signal averaged for 20 s, the background spectrum of nitrogen for 20 s, and the cell was vented for 10 s between each gas insertion. The NIR results are compared with the GC results for a 16-h run in Figure 21.12, where the mean NIR-BTU value for every 20 min is compared with the 2h GC values. The NIR results agreed well with the GC values.

This NIR-BTU monitoring system was operated daily for a period of 4 weeks. Every morning the system calibration was checked with a sample of a gas standard. No changes had to be made to the calibration model over the duration of the investigation. In reality, the monitor could have been run unattended for the 4-week period.

21.6 NIR GAS-PHASE LIBRARY AND SEARCH WITH MIXTURES

21.6.1 PREPARATION OF THE SPECTRAL LIBRARY

A 100 compound library of NIR gas-phase spectra and 17 mixtures containing two, three, and four components of compounds included in the library were measured on a Digilab FTS-40N FTIR from 9000 to 3500 cm^{-1} at a resolution of 2 cm^{-1} by co-adding 1024 scans [21]. The gas samples were

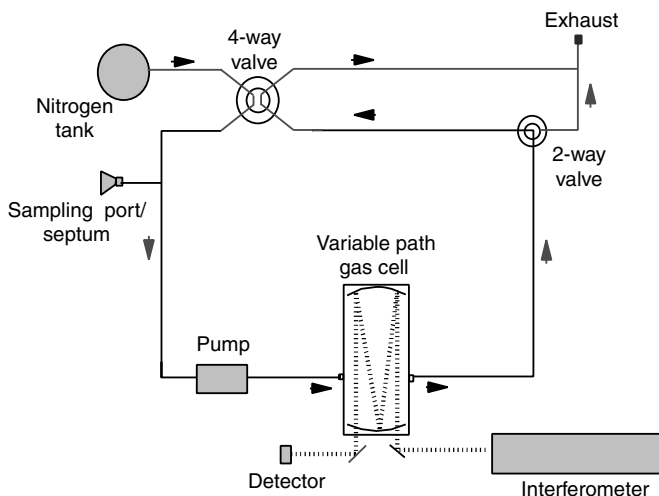


FIGURE 21.13 Schematic of flow-through gas system and 20-m variable path gas cell.

contained in a Foxboro Corp. (Foxboro, MA) 20-m variable-path gas cell and the measurements were performed at $20 \pm 0.5^\circ\text{C}$. Samples were introduced into and flushed from the cell via a gas-handling system built in-house. A diagram of the gas-handling system is shown in Figure 21.13. For gas mixtures, the individual components could be introduced at the septum and continuously cycled through the system to facilitate mixing with N_2 gas. Gases could be sampled from any external source and either pumped through or cycled through the variable-path gas cell. Gases could be vented or pumped to an exhaust.

21.6.2 SEARCH ALGORITHM

The algorithm used for searching the spectral library of pure compounds with spectra of mixtures is referred to as Mix-Match [22–30]. A flow diagram of the algorithm is shown in Figure 21.14. An orthonormal basis set of Principal Component (PC) spectra is generated from the library spectra using Principal Component Analysis (PCA). Both the library spectra and the unknown spectrum are projected onto this basis set to produce a set of scores, which are simply the coefficients used to regenerate each of the original spectra as a linear combination of the PC spectra. It is assumed that each of the library spectra represents a pure component having a concentration of 1.0; these pseudo-concentrations are regressed onto the scores of the pure components to obtain a regression matrix. Scores for the unknown spectrum are multiplied by the regression matrix to predict concentrations (referred to as the composition index or CI) of the library components.

To understand this process, consider as an example that an unknown is a 50/50 mixture of library sample #6 and #37. Under ideal conditions the algorithm would predict that the CI for the unknown was 0.5 for #6 and 0.5 for #37; all of the other composition indices would be close to 0.0. In real unknown mixture spectra, the prediction is not this definitive, but it works well considering its simplicity.

Normally, library spectra are not measured quantitatively and it is not possible to determine the absorptivity spectra, that is, spectrum of that compound at unit concentration. Instead, the library spectra are simply normalized so that the total area is 1.0. Thus, when an unknown 50/50 mixture is searched rarely will the results be 50/50. Typically, the match for one of the components will be greater than the match for the other component, for example, the composition index might be 0.65 for one and 0.30 for the other. The composition indices will rarely add to 1.0 owing to differences in the absorptivity spectra for the components. Moreover, there will be other spectra in the library that have a certain degree of similarity with the mixture spectrum, so that the composition indices will be >0 for a number of the spectra. There is an added possibility that the first search with the

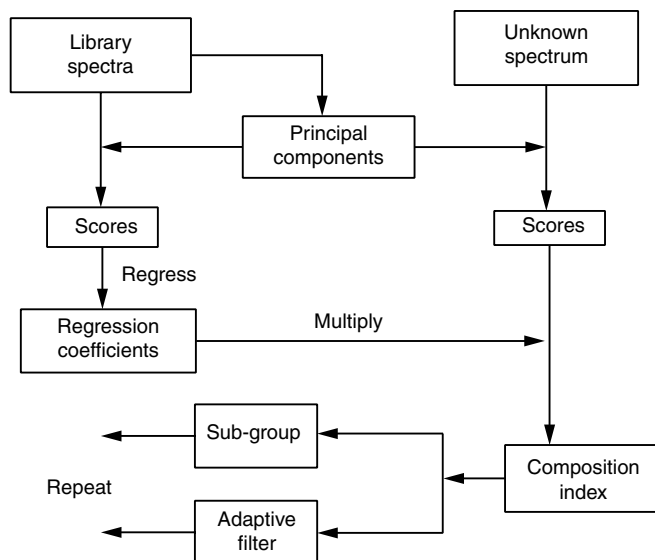


FIGURE 21.14 Flow-chart of the Mix-Match search algorithm.

Mix-Match algorithm produces a number of possible hits for the components in the mixture without any single hit having a large, definitive CI. This could happen if there were a number of library compounds with functional groups similar to those of one of the components in the mixture.

The Mix-Match algorithm provides two options for reprocessing the spectra after the first pass. In the first option, a subgroup of the library spectra can be selected and processed with PCA to provide a better fit to the unknown spectrum. Typically, the top 20 hits from the first pass are selected for the subgroup. Generally, the pass through the subgroup will produce a much better match with the target spectrum. This option is very useful when several library spectra produced similar CI values during the first pass.

The second option for reprocessing the library spectra is referred to as the Adaptive Filter method. This filter is used when a clear hit with a high CI is obtained during the first or second pass through the library. All of the library spectra and the spectrum of the unknown mixture are made orthogonal to library spectrum having the high CI, that is, all similarity of this spectrum is removed from the library spectra and from the unknown spectrum. These modified library spectra are subjected to PCA to produce a new basis set of orthonormal PC spectra. The scores for the modified library spectra and for the modified unknown spectrum are determined and a new list of composition indices is produced. The adaptive filter can be applied until all of the residual information in the unknown mixture spectrum has been used, that is, until the composition indices are <0.1 .

For the gas-phase spectra considered herein, one of the components in the mixture spectra was always identified on the first pass through the library and a second pass with a small subgroup was never necessary. The subgroup option is more important for spectra of liquids and solutions since the bands are broader and there is considerably more overlap. However, for gases, the bands are sharp and the identification is clearer.

21.6.3 SEARCH RESULTS

A total of 17 mixture spectra were searched in the 100 compound library. Compositions for the 17 mixtures are listed in Table 21.3. All of the library spectra were measured at 2 cm^{-1} resolution. The first 13 unknown mixture spectra were also measured at 2 cm^{-1} resolution. The two-component

TABLE 21.3
Results of Mixture Search on 17 Mixtures

Mixture	Components	Passes through library				
		1	2	3	4	5
Mix1	Benzene	2	1			
	Chloroform	1	—			
Mix2	<i>p</i> -Xylene	2	1			
	<i>p</i> -Chlorotoluene	1	—			
Mix3	Chlorobenzene	1	—			
	Benzaldehyde	2	1			
Mix4	1,3-Dichloropropane	1	—			
	2-Bromobutane	3	1			
Mix5	DMSO	1	—			
	THF	2	1	—		
Mix6	1,3-Dichloropropane	2	2	1		
	1,2-Dibromoethane	4	1	—		
	1,1,1,2-Tetrachloroethane	1	—	—		
Mix7	1,1-Dichloroethane	1	—	—		
	Chlorobenzene	3	1	—		
	Cyclopentanone	2	2	1		
Mix8	3-Heptyn	2	1	—		
	<i>n</i> -Hexane	4	3	1		
	1,1,2-Trichloroethene	1	—	—		
Mix9	<i>p</i> -Xylene	1	—	—		
	Toluene	3	2	1		
	1,2,4-Trimethylbenzene	2	1			
Mix10	<i>m</i> -Xylene	1	—	—		
	<i>p</i> -Xylene	2	1	—		
	<i>o</i> -Xylene	3	2	1		
Mix11	Cyclopentanone	1	—	—	—	
	1,2-Dibromoethane	2	1	—	—	
	<i>m</i> -Xylene	5	4	1	—	
	2-Methyl-pentane	6	3	2	1	—
Mix12	Acetyl chloride	2	2	1		
	<i>o</i> -Chlorotoluene	3	1	—	—	
	Butyl sulfide	1			—	
	2-Hexanone	6	4	2	1	
Mix13	<i>p</i> -Xylene	1	—	—	—	
	Toluene	3	3	2	1	
	Chlorobenzene	5	1	—	—	
	Benzyl chloride	2	2	1	—	
Mix14	Chlorobenzene	1	—			
	1-Bromo-3-Chloro-propane	2	1			
Mix15	1-Bromo-3-Chloro-propane	1	—			
	2-Pentanone	5	1			
Mix16	2,2-Dimethylbutane	2	1			
	3-Heptyn	1	—			
Mix17	Methanol	1	—			
	Cyclohexane	2	1			

Spectra for Mixtures 14 and 15 were measured at 8 cm⁻¹.

Spectra for Mixtures 16 and 17 were measured at 16 cm⁻¹.

mixtures #14 and #15 were measured at 8 cm^{-1} resolution; the two-component mixtures #16 and #17 were measured at 16 cm^{-1} resolution. Components in the 17 test mixtures were selected to include compounds representing most of the typical chemical functional groups. A number of the mixtures contained similar compounds such as all aromatics, since our goal was to demonstrate that NIR spectra of gases could be effectively used for detection and identification of mixtures of similar compounds. For example, Mix2, Mix3, Mix9, Mix10, and Mix13 contained only substituted aromatic compounds, whereas Mix6 contained only halogenated alkanes. Other mixtures contained both aromatic and nonaromatic compounds to simulate potential interferences that might be found in real samples.

Three examples of the search results will be given to demonstrate the potential of the NIR region for monitoring gas mixtures. The spectra of the mixture and the pure components along with the search results for the two-component mixture #5 containing approximately equal amounts of dimethylsulfoxide (DMSO) and tetrahydrofuran (THF) are shown in Figure 21.15. The last row of

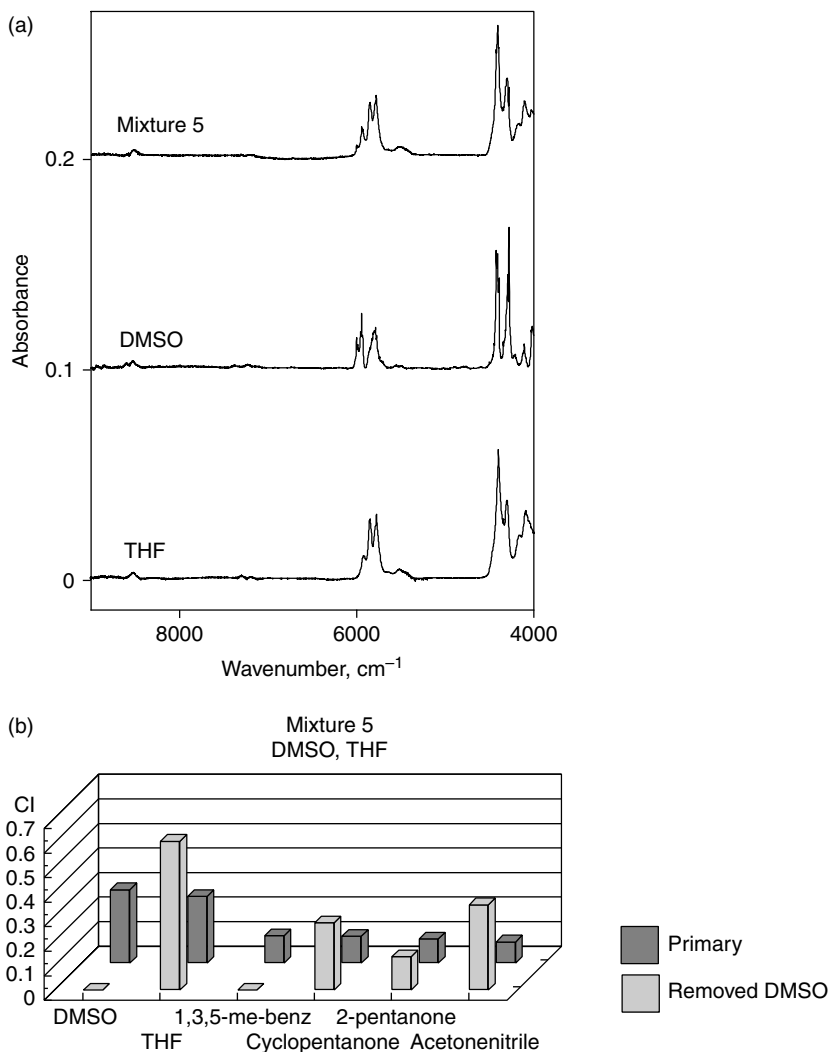


FIGURE 21.15 (a) Spectra of unknown two-component mixture and spectra of the pure components. (b) Bar-graph showing results of the Mix-Match search.

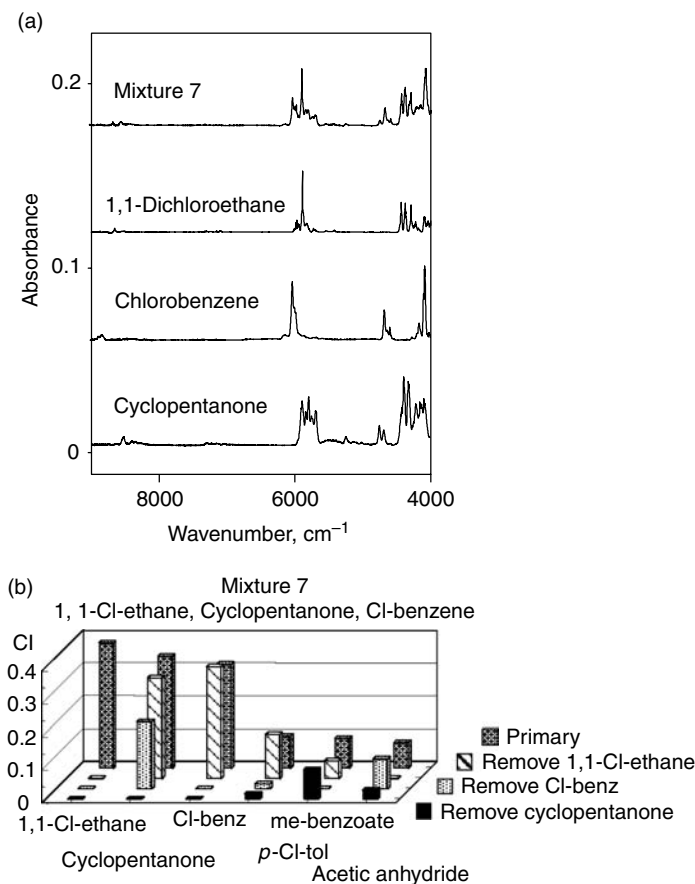


FIGURE 21.16 (a) Spectra of unknown three-component mixture and spectra of the pure components. (b) Bar-graph showing results of the Mix-Match search.

bars (solid black) represents the first pass through the library; DMSO was the #1 hit with a CI of 0.26 and THF was the #2 hit with a CI of 0.22. In the second pass, the library spectra and the unknown mixture spectrum were made orthogonal to the library spectrum of DMSO. The PCA processing was repeated on the modified library spectra and the algorithm applied to the modified spectrum of the mixture. THF was clearly the #1 hit on the second pass with a CI > 0.6. The modified library spectra and mixture spectra were made orthogonal to the THF spectrum and the process repeated. The final results from the third pass showed that all of the residual information has been removed and the identification is complete.

Spectra and bar-graph results for a three-component mixture of 1,1-dichloroethane, chlorobenzene, and cyclopentanone along with the search results are shown in Figure 21.16. In the first pass, 1,1-dichloroethane was the first hit, cyclopentanone was second, and chlorobenzene was third. The library spectra and the unknown were made orthogonal to the spectra of 1,1-dichloroethane and the process repeated to pick chlorobenzene as first and cyclopentanone as second. The modified spectra were made orthogonal to the spectrum of chlorobenzene and the process repeated to pick cyclopentanone as the top hit.

A rather difficult four-component example is shown in Figure 21.17. The bar-graph results at the bottom of the figure show that three of the components, butyl sulfide, acetyl chloride and *o*-chlorotolulene were the top three hits but the fourth component, 2-hexanone was the sixth hit. On the second pass, *o*-chlorotolulene was the top hit followed by acetyl chloride. On the third pass,

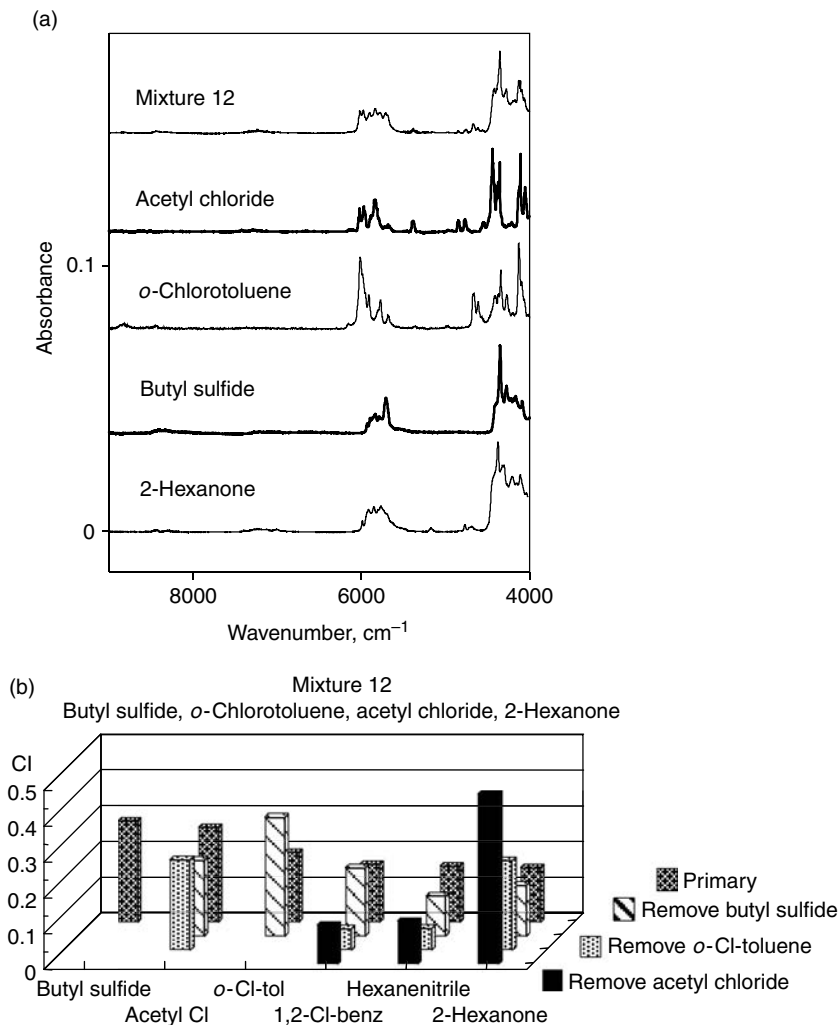


FIGURE 21.17 (a) Spectra of unknown four-component mixture and spectra of the pure components. (b) Bar-graph showing results of the Mix-Match search.

acetyl chloride was the top hit followed by 2-hexanone. Finally, on the fourth pass, 2-hexanone became the top hit. Thus, all four components were correctly identified.

21.6.4 SUMMARY OF THE SEARCH RESULTS

The initial searches for the 17 unknown test mixture spectra were performed using the 4000 to 8000 cm^{-1} spectral region. All of the components in all of the mixtures were correctly identified except for *o*-xylene in the three-component Mix10 and 2-hexanone in the four-component Mix12. The search was repeated using only the 4000 to 6200 cm^{-1} spectral region. During the first pass through the library, all of the anticipated components were among the top six hits; however, one or more of the components were in the fifth or sixth position. After the Adaptive Filter was applied one or more times, each of the analytes in all of the mixtures was correctly identified.

The spectral data for the mixture search can be preprocessed so that the actual search is very rapid. On a 2-GHz PC, a single pass through the library can be accomplished in less than 5 s.

A four-component mixture can be identified using repetitive applications of the adaptive filtering in less than 1 min. Certainly, the data content of NIR spectra is sufficient for making positive identifications even with mixture-type interferences.

21.7 CONCLUSIONS

At the beginning of this chapter, we stated that NIR has not been used extensively for monitoring and investigating gases. There are many advantages of using NIR for gases, not the least of which is the availability of detectors for the region. Absorptivities are weaker in the NIR compared to those in the MIR, but the potential for better signal-to-noise optical systems seems to compensate for the weaker absorptivities. We have shown herein that spectral resolution is not a deterrent to either quantitative or qualitative studies. Eventually, NIR will find its way into the gas-phase world and be just as successful as it has been for monitoring food and pharmaceutical products.

REFERENCES

1. W.F. McClure, *Anal. Chem.*, **66**: 46A–53A (1994).
2. W.F. McClure, in *Spectroscopic Techniques in Food Analysis*, R.H. Wilson, Ed., VCH New York (1993), pp. 13–57.
3. R.H. Wilson and B.J. Goodfellow, in *Spectroscopic Techniques in Food Analysis*, R.H. Wilson, Ed., VCH New York (1993), pp. 59–85.
4. M.A. Dempster, B.F. MacDonald, P.J. Gemperline, and N.R. Boyer, *Anal. Chim Acta*, **310**: 1005–1008 (1995).
5. M.G. Hansen and A. Khettry, *Polym. Eng. Sci.*, **34**: 1758–1766 (1994).
6. W.T. Welch, M.L. Bain, S.M. Maggard, and J.M. May, *Oil Gas J.*, **92**: 48–56 (1994).
7. J. Workman, *J. Near Infrared Spectrosc.*, **1**: 221–245 (1993).
8. J. Workman, *J. Near Infrared Spectrosc.*, **4**: 69–74 (1996).
9. M.L. McKelvy, T.R. Britt, B.L. Davis, et al., *Anal. Chem.*, **68**: 93R–160R (1996).
10. T.L. Marshall, C.T. Chaffin, R.M. Hammaker, and W. Fateley, *Environ. Sci. Tech.*, **28**: 224A–232A (1994).
11. D. Richter, D.G. Lancaster, and F.K. Tittel, *Appl. Optics*, **39**: 4444–4450 (2000).
12. J.A. Silver and D.C. Hovde, *J. Atm. Ocean Tech.*, **15**: 29–36 (1996).
13. D.B. Oh, M.E. Paige, and D.S. Bomse, *Appl. Optics*, **37**: 2499–2501 (1998).
14. Z. Yang, G.C. Toon, J.S. Margolis, and P.O. Wennberg, *Geophys. Res. Lett.*, **29**: 53-1-4 (2002).
15. K. Chan, H. Ito, and H. Inaba, *J. Lightwave Tech.*, **LT-2**: 234–237 (1984).
16. K. Chan, H. Ito, and H. Inaba, *J. Lightwave Tech.*, **5**: 1706–1710 (1987).
17. M.C. Alacon, H. Ito, and H. Inaba, *Appl. Phys. B*, **43**: 79–83 (1987).
18. R.T. Ellington, Department of Chemical Engineer, University of Oklahoma.
19. Headwall Photonics, (formerly American Holographics, Inc), Fitchburg, MA, USA.
20. C.W. Brown and J. Zhou, *Appl. Spectrosc.*, **55**: 44–49 (2001).
21. C.W. Brown, Optical BTU Sensor Development Final Report, Gas Research Institute, Chicago, IL, (1993).
22. C.W. Brown and S.-C. Lo, *Appl. Spectrosc.*, **47**: 812–815 (1993).
23. S.-C. Lo and C.W. Brown, *Appl. Spectrosc.*, **45**: 1621–1627 (1991).
24. S.-C. Lo and C.W. Brown, *Appl. Spectrosc.*, **45**: 1628–1632 (1991).
25. S.-C. Lo and C.W. Brown, *Appl. Spectrosc.*, **46**: 790–796 (1992).
26. C.W. Brown, A.E. Okefor, S.M. Donahue, and S.-C. Lo, *Appl. Spectrosc.*, **49**: 1022–1027 (1995).
27. J. Lin, J. Zhou, and C.W. Brown, *Appl. Spectrosc.*, **50**: 444–448 (1996).
28. C.-S. Chen, Y. Li, and C.W. Brown, *Vibrat. Spect.*, **14**: 9–17 (1997).
29. C.-S. Chen, C.W. Brown, and M.J. Bide, *J. Soc. Dyers Colourists*, **113**: 51–56 (1997).
30. C.W. Brown and S.-C. Lo, *Anal. Chem.*, **70**: 2983–2990 (1998).

22 Application for NIR Analysis of Beverages

Lana R. Kington and Tom M. Jones

CONTENTS

22.1	Introduction	457
22.1.1	Sample Preparation	457
22.1.2	Liquid Cell Design	457
22.2	Alcoholic Beverages	458
22.2.1	Beers	458
22.2.2	Wines and Distilled Liquors	458
22.3	Nonalcoholic Beverages	459
22.3.1	Fruit Juices	459
22.3.2	Glucose in Water Solutions	460
22.4	Formula Products	460
22.4.1	Historical Background	461
22.4.2	Accuracy and Precision	461
22.5	Conclusions	462
	References	463

22.1 INTRODUCTION

Important applications of near-infrared analysis (NIRA) are found in the beverage and nutritional formula industries. Technologies have been described for near-infrared (NIR) determinations of constituents in alcoholic beverages such as beer, wines, and distilled spirits; nonalcoholic beverages such as fruit juices, coffees, teas, and soft drinks; and other products such as infant and adult nutritional formulas. Some of these applications are described in this chapter.

22.1.1 SAMPLE PREPARATION

Liquid NIRA differs from NIR applications for ground or powdered samples in a few key ways. Rather than utilizing true reflectance measurements, liquid analysis involves transmission or a combination of transmission and reflectance sometimes referred to as transreflectance [1]. Another way in which liquid measurement differs from the measurement of powders involves the introduction of the sample to the instrument. Preparation of a liquid sample may involve dilution, suspension, or homogenization prior to sample introduction. A different type of sample cell, usually a flow cell of some type, is then required.

22.1.2 LIQUID CELL DESIGN

Several general factors should be weighed when designing or choosing a liquid cell. The mode of detection-transmission or transreflectance will influence the decision. Other factors to consider are the

physical characteristics of the sample. A liquid sample could be a slurry, a suspension, a solution, or an emulsion. An additional consideration is the application environment in which the sample will be used. Remote applications could utilize a flow cell in which the sample is injected manually or is processed via a homogenizer and pumped into the cell automatically. Online applications would probably require the ability for the sample to flow continuously or alternately fill and empty the cell as needed during process control.

Liquid cell design often takes into account the effect of temperature on reproducibility of measurements. Temperature effects on NIR protein are reported to be a factor in differences discovered between Federal Grain Inspection Service Laboratories in the early 1980s [2]. Williams et al. in a 1983 study conclude that an inverse relationship exists between temperatures of ground samples and their corresponding apparent protein contents [3]. Three suggestions for compensating for the differences are presented. One method is to select wavelengths that are minimally sensitive to temperature fluctuations. A second approach is to analyze samples at the same temperature at which the calibration was developed. The third suggestion is to develop a temperature compensation factor by use of a correlation chart. This sensitivity to temperature has also been noted by users of NIR technology for liquid analysis [4]. Although the preferred tactic for addressing this potential problem is to select wavelengths that are not temperature sensitive, many available liquid cells are designed with mechanisms to maintain constant temperature.

22.2 ALCOHOLIC BEVERAGES

NIRA is widely utilized for the analysis of constituents in a variety of alcoholic beverages. These include beer, wine, and distilled spirits. Applications range from direct readings for alcohol content to calibrations for original gravity, a function of alcohol.

22.2.1 BEERS

NIRA has been applied for determinations of alcohol and other constituents in beers. Coventry and Hunston have reported calibrations for alcohol in beers using transmission [5]. The most successful calibration for alcohol in a variety of light and dark beers utilized a single wavelength, 1672 nm. Standard errors of the calibration ranged from 0.2 to 0.3 over a range of 0 to 11% alcohol content.

The European brewing industry uses NIRA extensively to monitor the brewing process. Critical parameters measured by NIRA include alcohol; original gravity which is a function of alcohol; real extract; apparent extract; sugars; and total solids. To ensure reproducibility of these measurements, attention is given maintaining a fixed sample temperature. One method of achieving this result is circulate water at the desired temperature through a modified solid sample compartment. A cell for this purpose has been developed by Technicon Europe. Approximately 20 ml of liquid is injected into the cell. Equilibration time to achieve the desired temperature is only about 30 s. Some key wavelengths for these types of calibrations described by Dzwinczyk are presented in Table 22.1 [6].

In a related application, the brewing industry has also extended the uses for NIRA to include analysis of raw material components. Total nitrogen in barley and malt and α -acids in hops are two such applications [7].

22.2.2 WINES AND DISTILLED LIQUORS

In routine use in more than 50 sites in France is the NIR measurement of alcohol and sugars in wine and distilled liquors [7]. The concentration of alcohol in wine was determined by Kaffka and Norris as early as 1976 [8]. The accuracy of their calibration is reported as better than 0.1% by volume at 11 to 17% alcohol using transmittance as the mode of detection. Sucrose in wine has also been investigated. An accuracy of 0.2% is achieved with a calibration for sucrose levels of 0.8 to 8.8%.

TABLE 22.1
Wavelengths Used in Calibrations
to Determine Composition of Beers

Constituent	Wavelength (nm)
Alcohol (ref.)	2230
Alcohol	2310
Water	1445
Sugar	2100

Source: M. Dzwinczyk, *Proc. 8th Int. Symposium on Near Infrared Reflectance Analysis (NIRA)*, Technicon, Tarrytown, New York, 1985.

TABLE 22.2
NIR Calibrations for Brix and Acid over Entire Range

	Brix (Range = 4.4–19%)	Acid (Range = 0.3–2.26%)
Wavelengths (cm)	1445	1828
	1778	2010
	1818	1576
	2190	2150
<i>R</i>	0.999	0.996
SEE	0.200	0.041
<i>N</i>	66	66
<i>F</i> ratio	7348	1779

Source: Courtesy of L. P. McDermott, Technicon Industrial Systems, Technicon, Tarrytown, New York, personal communication, 1988.

The same authors reported a determination of tartaric acid, another constituent of interest in the wine-making industry, over a range of 0.5 to 8.2% with an accuracy of 0.1%.

22.3 NONALCOHOLIC BEVERAGES

Constituents present in many nonalcoholic drinks have also been quantitated with NIRA, for example, moisture and caffeine in coffee [7]. Other applications have been developed for constituents in teas, soft drinks, and fruit juices. Most of these applications are unpublished.

22.3.1 FRUIT JUICES

Calibrations for degrees Brix and acidity in fruit juices have been performed utilizing transfectance and a 45°C liquid cell [9]. The calibration parameters are presented in Table 22.2 to Table 22.4 with the wavelengths utilized for each calibration.

Carbohydrate in fruit juices has been determined using NIRA transmittance with a 2.2-nm path-length cell [10]. Such calibrations have worked best when performed for each type of fruit, and have achieved relative standard deviations of 0.25% in the range of 9.3 to 11.3% for orange juice.

TABLE 22.3
NIR Calibrations for Brix on Basis of Percent Brix

	Brix range (%)		
	4–8	9–14.4	14.5–19
Wavelengths (nm)	1436	1352	1436
	1450	1744	2206
	1506	2220	2388
<i>R</i>	.990	.995	.997
SEE	0.161	0.146	0.127
<i>N</i>	15	29	22
<i>F</i> ratio	184	89	1072

Source: Courtesy of L. P. McDermott, Technicon Industrial Systems, Technicon, Tarrytown, New York, personal communication, 1988.

TABLE 22.4
NIR Calibrations for Acid on Basis of Percent Brix

	Brix range (%)		
	4–8	9–14.4	14.5–19
Wavelengths (nm)	2276	1436	1352
	1954	1828	1534
	2052	1758	1800
<i>R</i>	.999	.990	.960
SEE	0.023	0.038	0.024
<i>N</i>	15	29	22
<i>F</i> ratio	4822	405	70

Source: Courtesy of L. P. McDermott, Technicon Industrial Systems, Technicon, Tarrytown, New York, personal communication, 1988.

22.3.2 GLUCOSE IN WATER SOLUTIONS

NIRA calibrations have also been developed for glucose solutions manufactured for pediatric use [4]. These solutions typically range from 5 to 10% glucose by volume in water. The simplicity of preparing standard solutions directly from glucose raw material sources makes this an extremely rapid, accurate, and reproducible alternative to standard AOAC methodology.

22.4 FORMULA PRODUCTS

Infant formula products and other nutritional formulas directed to the needs of hospital patients provide a multibillion dollar annual market for several manufacturers. These liquid formulas are complex suspensions of nutrients, the levels of which are carefully controlled. Regulation of these products, especially infant formulas, dictates the need for a complex array of nutrient analyses.

TABLE 22.5
Soy Formula Calibration for Major Constituents

	Protein	Fat	Solids
Wavelengths (nm)	2100 2180	1759 2270 2310 2348	1818 1982 2139
R^2	.9933	.9989	.9998
SEE	0.0770	0.0529	0.0751
N	31	31	31
F ratio	6316	19378	172592

Source: Courtesy of Bristol-Myers Squibb MJNG, T. M. Jones and L. R. Rademacher, *Proc. Rocky Mountain Conference*, Denver, Colorado, 1986.

While these preparations may often resemble milk or dairy products, they are actually quite different, and provide unique analytical problems. Several of the required tests, however, lend themselves well to spectroscopic techniques including NIRA.

22.4.1 HISTORICAL BACKGROUND

The use of NIRA for monitoring constituent levels in infant formula products began in the late 1970s. Protein, fat, content, and total solids monitoring were the targets of initial efforts in this area [11]. The early work paralleled the development of NIRA for dairy applications. Instrumentation was relatively simple at this point, utilizing a flow cell in the place of the powder sample holder. Samples were prewarmed and injected into the flow cell manually via syringe.

In the early 1980s, significant progress was made to expand the application of NIRA to other types of formula products. Calibrations for constituents in several soy-based, milk-based, and glucose/water formulations were developed [11,12].

These included infant- and hospital-directed products. Concurrently, instrumentation was greatly improved for dairy and liquid applications [13]. Samples were now preheated, homogenized, and introduced to the instrumentation through use of an external sample preparation module.

The evolution of formula applications involving NIRA extended to the process environment in 1986. Calibrations were developed for different process lines in which the protein, fat, and total solids levels were monitored at different process stages by laboratory personnel using benchtop equipment [14]. Wavelengths and regression statistics reported for these calibrations are produced in Table 22.5 and Table 22.6. The next logical extension of NIRA development would be to perform this testing using online techniques.

22.4.2 ACCURACY AND PRECISION

Utilizing NIRA for formula analysis has proven to be cost-effective and reliable [14]. The use of a careful check sample program ensures the continued accuracy and reliability of analytical data obtained by the technique over a long period of time [15]. Table 22.7 provides an indication of the accuracy and precision typical of formula applications.

A summary of specific liquid formula applications in which NIRA is used as a primary method of analysis is given in Table 22.8. As NIRA becomes more widely used in this industry, the list of applications will continue to expand.

TABLE 22.6
Milk-Base Formula Calibration for Major Constituents

	Protein	Fat	Solids
Wavelengths (nm)	1818	2270	1445
	2100	2310	2100
	2139	2336	1982
	2180		2270
R^2	.9946	.9993	.9995
SEE	0.535	0.0477	0.1446
N	45	45	45
F ratio	5977	62364	59852

Source: Courtesy of Bristol-Myers Squibb MJNG, T. M. Jones and L. R. Rademacher, *Proc. Rocky Mountain Conference*, Denver, Colorado, 1986.

TABLE 22.7
Typical Accuracy and Precision of NIRA for Liquid Formula Analysis

	Total solids	Fat/Oil	Protein
Precision (% RSD)	<0.5	<0.3	<0.5
Relative accuracy (%)	± 1	± 1	± 3
N	6	6	6

Source: Courtesy of Bristol-Myers Squibb MJNG, L. R. Rademacher, *Proc. 4th Int. Symposium on Near Infrared Reflectance Analysis*, Technicon, Tarrytown, New York, 1983.

TABLE 22.8
Liquid Formula NIRA Applications

Application	Constituents
Whey-casein infant formula	Protein, fat, solids
Soy protein infant formula	Protein, fat, solids
Glucose/water solution (5 and 10%)	Glucose
Complete isotonic formula	Protein, fat, solids
Complete liquid diet	Protein, fat, solids

Source: Courtesy of Bristol-Myers Squibb MJNG.

22.5 CONCLUSIONS

The use of NIRA for beverage and infant formula applications is rapidly increasing. Typically involving high-volume markets with concurrently large sample volumes in quality control, these product types are ideal candidates for NIR technologies. Investments in instrumentation, calibration

software, and operator training are easily and quickly recoverable. Relatively few NIR applications for beverage and formula products have been published to date, however, due to the highly competitive markets involved. This situation may well be resolved as the number of applications and participating laboratories continues to increase.

REFERENCES

1. D. A. Burns, NIR Resources, Putnam Valley, New York, personal communication, 1984.
2. Anon., *Mill. Bak. News*, 20: 33 (1981).
3. P. C. Williams, K. H. Norris, and W. S. Zarowski, *Cereal Chem.*, 59: 473 (1982).
4. N. U. Siddiqui and T. M. Jones, Bristol-Myers, Evansville, Indiana, unpublished study, 1981.
5. A. G. Coventry and M. J. Hunston, *Cereal Foods World*, 29: 715 (1984).
6. M. Dzwinczyk, NIRA in the Brewing Industry, *Proc. 8th Intl Symp. on Near Infrared Reflectance Analysis (NIRA)*, Technicon, Tarrytown, New York, 1985.
7. M. Day, *Food Proc.*, 53: 51 (1984).
8. K. J. Kaffka and K. H. Norris, *Acta Aliment.*, 5: 267 (1976).
9. L. P. McDermott, Technicon, Tarrytown, New York, personal communication, 1988.
10. B. G. Osborne and T. Fearn, *Near Infrared Spectroscopy in Food Analysis*, John Wiley & Sons, New York, 1986, p. 132.
11. N. U. Siddiqui, Bristol-Myers, Evansville, Indiana, personal communication, 1987.
12. T. M. Jones, Smith-Kline and French Laboratories, King of Prussia, Pennsylvania, personal communication, 1987.
13. K. Luchter, InfraAlyzer 400D Evaluation, *Proc. 4th Intl Symp. on Near Infrared Reflectance Analysis*, Technicon, Tarrytown, New York, 1983.
14. T. M. Jones and L. R. Rademacher, A Strategy for the Use of NIRA in Liquid Formula Quality Control Applications, *Proc. Rocky Mountain Conference*, Denver, Colorado, 1986.
15. L. R. Rademacher, Quantitative Aspects of NIRA: Its Value as a Control Tool, *Proc. 4th Intl Symp. on Near Infrared Reflectance Analysis*, Technicon, Tarrytown, New York, 1983.

23 NIR Analysis of Wool

Michael J. Hammersley and Patricia E. Townsend

CONTENTS

23.1 Introduction	465
23.2 Instrumentation	468
23.3 Calibration Set Assembly	470
23.4 Reference Methods	470
23.5 Calibration Mathematics	472
23.6 Extensions	477
References	478

23.1 INTRODUCTION

Both the wool production and the wool textile-manufacturing industries offer great scope for the application of near-infrared (NIR) techniques. The demand for raw (unprocessed) fiber testing is considerable; for commerce as well as to assist sheep-breeding programs. Applications as a quality control tool during textile manufacture would bring the benefits already obvious in those industries where NIR is now commonplace.

Even now the field is relatively unexploited. There is one major application where NIR is routinely currently in use, and providing an appreciable and appreciated benefit to its users: the analysis of washed wools in New Zealand, the world's major exporter of such wools. The success of that work has encouraged other, related applications but, as yet, these operate on a smaller scale. Most attention has been given to the raw wool area where perhaps the largest benefits might accrue in the short term but where the material exhibits its greatest variability — a marked emphasis on sampling and sample presentation is characteristic of raw wool applications.

Wool as shorn consists of a virtually pure protein fiber carrying two classes of undesirable contaminant: those produced by the normal metabolic processes of the sheep and those derived from the environment. The contaminants produced by the sheep are wool wax and suint. Suint is dried sweat and consists mainly of organic and inorganic salts of potassium. Wool wax, the waxy secretion produced by the sebaceous glands, is a very complex material containing high molecular weight esters together with some free alcohols and acids. Another contaminant that may be present is dung.

The externally derived contaminants are mainly fragments of vegetation such as seeds, leaves, small twigs, etc. picked up as the sheep grazes. Mineral matter is also present. It comes from windblown dust or from direct contact of the fleece with bare ground.

In addition to these contaminants, all wool contains some moisture. Wool, even though apparently dry, may contain up to some 30% moisture, a typical figure being about 14%, depending on the relative humidity of the surrounding air. Suint, present only in unwashed wool, also contains some moisture owing to its hygroscopic nature.

The wool content of greasy wool may be as low as 40 to 50% and the cost of the fiber ensures that there is a great interest for commercial purposes in measuring the “yield” — one definition of

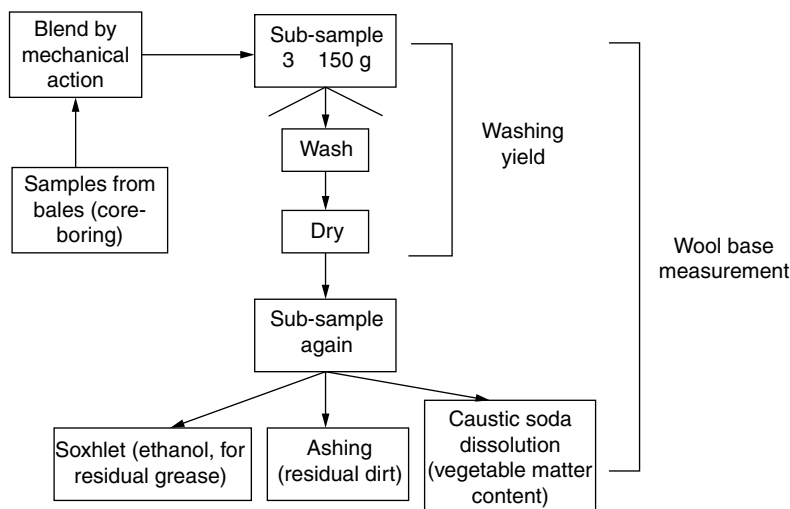


FIGURE 23.1 Summary of the wool yield tests.

the wool content. In the major wool-growing and exporting countries of Australia and New Zealand the annual costs of yield testing account for several million dollars.

The standard method for yield determination is published under the aegis of the International Wool Textile Organization (IWTO) [1]. A summary of the procedures is shown in Figure 23.1; the detailed description runs to 29 A5 pages. The earliest application of NIR to yield testing was reported by Sabbagh and Larsen [2]. Using a six-filter instrument (InfraAlyzer 2.5) they achieved a standard error of calibration (SEC) of 2.8% for wool base — the yield expressed as clean dry-fiber content. There were no validation experiments and the reported result refers to the SEC achieved with 25 wools ranging in wool base from 31 to 53%. In this pioneering work the problems of sampling and sample presentation were encountered for the first time. A full account will be found in Sabbagh's thesis [3].

These findings provoked a flurry of work in Australia. At the University of New South Wales, Scott and Roberts [4] started to use NIR to estimate clean fleece weight (CFW), the product of the greasy (as shorn) fleece weight and yield. CFW represents the quantity of wool grown since the previous shearing and it is used for sire (ram) selection. The yield measure used was the washing yield (see Figure 23.1), the traditional measure in this application. By this time the need for validation samples was recognized and furthermore the use of “mini-coring” methods for subsampling was evaluated. Mini-coring is analogous to the core-boring methods normally used to take samples from bales of wool. Core-boring cutters for bales may be about 20 mm in diameter, and Scott and Roberts' mini-cores were obtained with a cutter of 2 mm diameter.

Scott and Roberts' published data [4] provided no estimates of either SEC or the standard error of prediction (SEP) of unknowns. For CFW correlations with the reference results of .99, .99, and .96 are reported for three flocks, each flock having its own calibration equation. The correlations for CFW were slightly better than those for washing yield itself.

Scott and Roberts [4] also noted that they could estimate mean fiber diameter (MFD) from NIR data. This important parameter ranges from about 18 to 45 μm or so. To a large extent, fiber diameter indicates the end use and fine wools command higher prices. Because grinding procedures are inappropriate for wool, the sensitivity of NIR spectra to particle size, normally seen as unfortunate and undesirable, becomes an exploitable property and gives an estimate of MFD.

Later work on this aspect reported by Scott et al. [5] claimed correlations of 0.9 against one of the standard methods, the fiber-diameter-by-airflow method [6]. Unfortunately these encouraging results were obtained only with clean fiber; working with greasy fiber the correlation fell to .74.

Similar results for wool base estimates were reported by Larsen and Kinnison [7] using a six-filter InfraAlyzer 400 with samples of commercial lots of wool as well as ram fleece samples, and they too observed that high correlations with fiber diameter were obtainable. For fiber diameter estimates, correlation coefficients of .93 are quoted for the commercial lots in both the greasy and the washed states while a value of .94 was obtained with the ram fleeces. However, these statistics refer only to the calibration set and the technique was not fully validated and put to use.

It is fair to observe that NIR is not likely to be the method of choice if MFD is of prime interest. However, a moderately reliable estimate of MFD may be valuable if it is gained in the course of other measurements, as would be the case with ram-screening programs.

Degree of medullation is another physical property which, like fiber diameter, might be expected to be not predictable by NIR. Medullation is a characteristic seen mainly in the coarser wools; the fiber contains voids that may be connected so that the fiber forms a tube. It has significance for some textile products and its presence often needs to be detected and quantified. Ranford et al. [8] initially investigated the use of NIR as an alternative to the normal reference method, projection microscopy, which is very tedious. While the results were encouraging, the precision achieved was relatively poor owing mainly to the low precision of the reference results. Later experiments used medullation estimates obtained from the WRONZ medullameter [9], a device in which the wool substance is rendered transparent by immersion in a liquid of the same refractive index. The medulla are thereby made visible and the degree of medullation is measured photoelectrically. A significant improvement in SEP was produced by this change and, following some further refinements, medullation estimates were added in 1990 to the other standard parameters that might be measured in wool scours [10].

Another approach to cost reduction in yield testing was made by Slack-Smith et al. [11] at the Australian Wool Testing Authority where attempts were made to use NIR estimates of residual fatty matter in scoured (washed) wool as a replacement for the soxhlet extraction procedure of the standard yield test (Figure 23.1). A set of 32 scoured wool cores was used to calibrate Technicon InfraAlyzer and Neotec GQA 31 instruments via multiple regression techniques. Validation performance on 919 “unknowns” was considered satisfactory in that bias was largely absent and the between-test-specimen variances for the NIR results were similar to those obtained from the soxhlet method. Although an attempt was made to introduce the technique on a large scale, it was found that the calibration equations were not significantly robust and the method was abandoned.

Connell and his colleagues at the CSIRO Division of Textile Physics carried out a very wide-ranging investigation in the hope of using NIR to estimate yield. Their early work was dominated by an emphasis on calibration methods and very encouraging statistics were produced based on the calibration sets.

Australian wools may have vegetable matter (VM) contents which are high by New Zealand standards ($>1\%$). Connell and Brown [12] observed that it was useful to supplement NIR data with reflectance measurements in the visible region because the VM is usually darker than the wool. Using a six-filter InfraAlyzer together with a Macbeth Corporation Colour Eye they derived equations that were claimed to give a precision (95% CI) of $\pm 1.8\%$ for wool base that compares well with the $\pm 1\%$ of the standard method.

Connell's group continued their researches using spectrophotometers such as the Neotec 6350 rather than discrete filter instruments and Connell and Norris [13] claimed that a small improvement in performance could be obtained by using “non-standard” wavelengths when measuring washed wool. Support for this view was expressed later by Higgerson et al. [14] working with greasy wool and citing a reduction of SEC from 2.2 to 3.0 to 1.8% by using the nonstandard wavelengths available in the 19-filter InfraAlyzer 400R.

Similar figures for SEC were also quoted by Connell and Norris [15] and again by Connell [16] for greasy wool. Only in this latter paper was the matter of validation seriously considered. Unfortunately, the SEP values obtained were so poor that the research effort was stopped.

With hindsight it seems likely that at least part of the reason for this unfortunate outcome lay in the use of inadequate calibration sets.

In New Zealand the group at the Wool Research Organisation of New Zealand, Inc., (WRONZ) concentrated initially on the properties of scoured wool. Scouring is basically a wash in an aqueous detergent solution followed by rinsing and drying; there are accessory processes that may be performed such as the blending of wool lots, bleaching and/or the application of other chemical treatments. In contrast to the Australian case most of the New Zealand wool clip is exported scoured.

The parameters of primary interest are the levels of remaining wool wax, called residual grease, and the moisture content. For residual grease the reference method is soxhlet extraction with dichloromethane, which also removes detergent residues, usually minute. Normal residual grease figures run from about 0.15 to about 0.6% of the weight of clean, dry wool. An upper limit of 0.5% is commonly specified so as to avoid problems in processing the wool.

Moisture content is commonly expressed as a percentage of the weight of clean dry fiber and called regain. The reference method involves drying to constant weight at $105 \pm 2^\circ\text{C}$. Typical figures range from about 14% to about 20%. High regains incur the risk of mildew during storage or shipping.

NIR methods were first introduced to the scouring industry in the form of 19 filter InfraAlyzers providing measurements of residual grease and regain. As the hardware evolved so did the methods and measurements change. Thus the introduction of the Foss-NIRSystems Model 6500, with a spectral range from 400 to 2500 nm, allowed the introduction of color measurement. The very recent introduction of a diode array-based instrument, very fast and “viewing” a relatively large area, has allowed radical changes to sample handling to the extent that fully automatic unattended operation is feasible.

In short, the initial InfraAlyzer-based work, on residual grease in particular, both developed the WRONZ skills and ensured a secure place in the scouring industry for the technology. Hardware evolution has allowed expansion of the suite of measurements and a yet closer approach to the industry ideal — an online, fully automatic operation that measures everything of interest at very little cost. These developments in New Zealand represent the longest-lived, most extensive, and most fully evolved industrial application in the wool industry. This success has given impetus to similar developments for greasy wool and for partly processed materials. A description of the scoured wool implementation comprises most of the remainder of this chapter, with emphasis given to those aspects peculiar to it.

23.2 INSTRUMENTATION

Whilst the instrumentation from the established manufacturers almost invariably performs to specification, it suffers, inevitably perhaps, from being designed as “general purpose.” One chooses the sample holders most appropriate for one’s material and similarly the operational mode, the data treatment methods etc. To various degrees one fits the material to the machine rather than the other way around. For the first two generations of NIR spectrophotometers (InfraAlyzers, Model 6500s) deployed in the wool industry this was emphatically true. Only with the most recent developments, a joint effort between KES Analysis Inc. of New York and WRONZ, has the instrumentation been consciously designed to suit wool. An outline of the salient characteristics of the three instruments follows:

InfraAlyzer 400R:

19 filters, 1445 to 2348 nm

Area of sample seen:

36 cm² (after WRONZ modification)

Time to measure one test specimen:

approximately 5.25 min

Sample handling/presentation:

10 g test specimen obtained from ~100 g primary (core-bored) sample by mini-coring [17]. A “wool drawer” and associated sample cup designed

by WRONZ allows readings from eight nonoverlapping areas that increase the total area “seen” and slows down the instrument markedly. The use of eight readings is not excessive. The standard deviation of the mean of eight readings on wool is comparable with but still higher than that of a single reading on wheat flour, a product with which the InfraAlyzer has had considerable success.

Parameters measured:

residual grease, regain, medullation (in one installation only)[†]

Foss-NIRSystems Model 6500:

400 to 2500 nm with 2 nm resolution

Area of sample seen:

80 cm² using a WRONZ cell similar to the Foss-NIRSystems high fat/moisture cup.

Time to measure one test specimen:

approximately 1.7 min

Sample handling/presentation:

10 g test specimen obtained from ~100 g primary (core-bored) sample by hand

Parameters measured:

residual grease, regain, “as-is” color,* base color,* fiber diameter,[†] medullation[†]

KES/WRONZ NiraSpec:

400 to 1700 nm with 7.3 nm resolution from 400 to 1100 nm and 10 nm resolution from 1100 to 1700 nm

Area of sample seen:

260 cm²

Time to measure one test specimen:

approximately 1 min

Sample handling/presentation:

entire core-bored sample used as test specimen

Parameters measured:

residual grease, regain, “as-is” color,* base color,* fiber diameter, medullation[†]

* A description of these wool characteristics follows in Section 23.4 (reference methods).

[†] Morphological attribute, NIR estimates are desired.

Thus, the major advantages of the Model 6500 over the InfraAlyzer lay in the extended spectral range, thereby allowing color measurements, and in the larger area “seen” by the instrument, which allowed the elimination of mini-coring and higher throughput. That increase in throughput became essential; scour productivity has risen steadily and eventually InfraAlyzers could no longer keep up. Ironically, part of the reason for the increased productivity lies in the introduction of NIR technology.

With regard to color measurement, it is fair to observe that the Model 6500 was not designed as a colorimetric instrument. The optical geometry does not comply with specifications of the Commission International de L'Eclairage (CIE), there is no method of setting the zero of the reflectance scale, the white standard is not specified, and, in the authors' experience, it varies appreciably between instruments. Furthermore there is no provision in the software to calculate color values from the spectral data. To some extent these objections can be overcome by treating color values as though they were constituents but this yields a level of performance below the true capabilities of the instrument. This point is discussed further in the Calibration Mathematics Section.



FIGURE 23.2 The evolution of sample size: 10 g mini cored sub-sample for an InfraAlyzer compared with 100 g complete primary sample for NiraSpec.

The major advantages of the NiraSpec arise from the sheer speed of measurement together with the size of sample “seen.” The latter, in turn, arises from the provision, to WRONZ, of the spectrophotometric components, leaving WRONZ to design the sample handling/presentation. Use of the entire core sample eliminates the labor of subsampling, eliminates the risk of regain changes induced by sample manipulation, and renders fully automatic operation possible. A further advantage is that provision can be made to obtain color values in the conventional way rather than by correlation (treating color values as constituents).

Examples of test specimens are shown in Figure 23.2 while a prototype NiraSpec instrument in routine industrial use but prior to automation, is shown in Figure 23.3.

23.3 CALIBRATION SET ASSEMBLY

One well-known constraint on the quality of predictive equations is the need for a comprehensive coverage of all normal variables within the calibration set. Fortunately, WRONZ is a research association serving almost the entire New Zealand wool industry and the collection of samples from all over the country was not too difficult. Even so, the relative paucity of high-grease-content wools remained as a constraint on the residual grease calibration range that is now 0.15 to 0.8%, although one InfraAlyzer installation was successfully extended to 1.1%. The problem has become worse in recent years as the now-extensive use of NIR methods in New Zealand wool scours means that it is very rare for out-of-spec ($>0.5\%$ residual grease) wools to be produced. Some wools were deliberately underscoured at WRONZ in an attempt to remedy the deficiency but this proved unhelpful.

Deliberate variation of the regain, on the other hand, has been found very satisfactory, it being only necessary to expose the wool to a dry atmosphere over silica gel or a moist atmosphere over water for a few hours at room temperature. A calibration range of about 8 to 22% regain appears to be entirely satisfactory in practice.

One further constraint must be mentioned. Wool-scouring technology is continually changing in attempts to improve the product and/or increase efficiency. Changes in detergent type and the use of chemical treatments such as the mothproofer application during scouring are but two of the changes whose introduction may render a calibration equation unsatisfactory for the new product.

23.4 REFERENCE METHODS

Another well-known constraint on the quality of predictive equations is the quality of the results from the standard methods. The lessons learned in other industries are no less relevant here and an

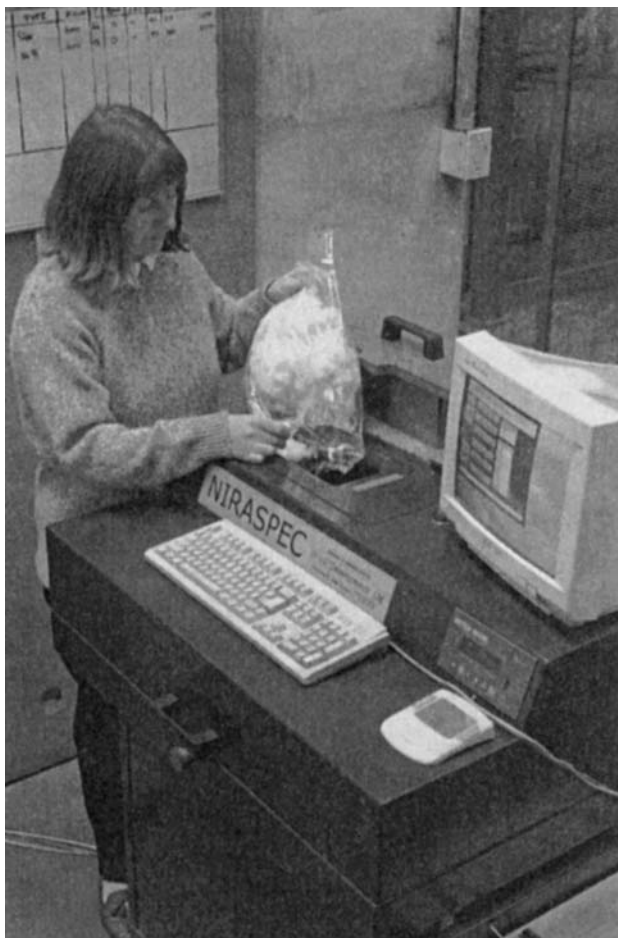


FIGURE 23.3 Hand loading a 100 g sample into NiraSpec during industrial use prior to automation.

intensive investigation of the nature of soxhlet extraction was necessary before it became possible to produce satisfactory NIR predictions of residual grease content.

Briefly, the conventional wisdom was that only the number of syphonings of the soxhlet apparatus was important, the time spent on the process being considered irrelevant. In fact, when dichloromethane is used as the solvent the reverse is true and reasonably close control of the process time is essential. As reported by Hammersley [18], once the procedure is appropriately controlled it is quite normal to achieve 95% confidence limits of $\pm 0.03\%$ extractable matter for the mean of two determinations within one laboratory even when using several operators. Comparable figures for soxhlet extractions performed prior to the establishment of the revised method ranged from ± 0.14 to $\pm 0.20\%$ extractable matter — this with a typical level of extractables of 0.3–0.4%!

In contrast, moisture content measurements were straightforward. The standard method involves drying the sample to constant weight with a forced draught of air at $105 \pm 2^\circ\text{C}$. Careful operatives are able to obtain easily the quality of result illustrated in Table 23.1. It might even be inferred from Table 23.1 that the NIR results may be more accurate as well as more precise than the reference results in that, provided bias is absent, the more precise result is usually also the more accurate result.

Color measurements are complicated by the industrial use of measurements made with the wool in two states. “As-is” color refers, not surprisingly, to measurements without modification to the material other than bringing it to the standard density for wool color measurements (160 kg/m^3).

Base color refers to measurements made after a further, rigorous cleaning of the wool [19]. Base color values are representative of the fiber alone, contributions to the color from surface impurities

TABLE 23.1
Reproducibility of Moisture Content
Measurements

Operator ^a	Reference		NIR	
	Mean	SD	Mean	SD
A	12.25	0.26	12.52	0.09
B	12.57	0.13	12.48	0.07
C	12.58	0.23	12.48	0.07
Overall	12.465	0.259	12.495	0.075

^a All operators used the same material. Each operator made ten measurements.

having been largely removed. Base color values are higher than the corresponding “as-is” values [20]. The difference is a valuable measure of the efficiency of the scouring and supplements the residual grease figure in that respect.

The IWTO reference methods for raw wool colorimetry [21] describe the precautions necessary when operating the colorimeter as well as the acceptable preparatory methods. Conventional commercially available colorimeters such as the Hunterlab ColorQuest and the ACS-Datcolor CS-3 are used. Instruments of 45/0 or 0/45 configuration are preferred but not essential.

“As-is” color measurements are not robust — the color can be altered very easily by the handling of the wool [20]. This characteristic makes it difficult to measure reproducibly; it also means that the innate value of the measurement is low. Nevertheless, within any one scouring plant, the difference between base tristimulus Y and “as-is” tristimulus Y is a valuable guide to scouring efficiency. This will be particularly so for the NiraSpec where the sample handling will be eventually automatic and therefore, presumably, reproducible.

The reference method adopted for “as-is” color measurements is essentially the same method used for the base color measurements [19] but without the rigorous cleaning procedures. Ideally no reference method (wool color measurement) is required for calibration but standard colorimetric measurements allow the calculation of “SEP” figures and hence some reassurance about the data quality.

Fiber diameter is measured by evaluating the resistance to the flow of air of a small mass of fibers prepared in a standardized way. The procedures are described by the IWTO [22]; it is a long-established technique used widely and giving little trouble in the majority of cases.

Medullation, when provided, is also based on an indirect method, in this case via the opacity resulting when the wool is immersed in a liquid of the same refractive index [9]. For both fiber diameter and medullation the primary reference method is projection microscopy. Such measurements are very slow, tedious and incompatible with the classic NIR requirements for large numbers of high-quality measurements. The use of the indirect methods effectively makes NIR calibration possible. It should be noted, however, that fiber diameter is best measured in other ways if the measurement is important and that the demand for medullation measurements is low. Neither of these morphological characteristics is affected by the scouring process.

23.5 CALIBRATION MATHEMATICS

The substantial changes in hardware of the successive generations induced major modifications to the sample handling/presentation methods, as already described. There were also changes to the calibration calculation procedures, initially because of the greater quantity of data.

It should be stressed at this point that the authors' stance with regard to validation populations is that all of their data should be collected after the calibration set data so as to most closely simulate what happens in practice. The common procedure of taking, for example, every third sample as a validation sample is not favored except, perhaps, as a screening exercise to clean up the calibration set. All of the WRONZ work reported here refers to such separate, post-calibration, validation sets.

InfraAlyzer 400 data, having a maximum of 19 data points/spectrum, are restricted in the variety of mathematical manipulations available for the purpose of generating a predictive equation. Even so, many methods were tried [23]. Eventually, the "branch and bound" technique of Furnival and Wilson [24] was found both satisfactory in performance and easy to apply, being part of the BMDP statistical software package [25]. With careful choice of the limiting parameters stable equations were consistently produced.

Moisture content equations typically used four or five variable $\log(1/R)$ terms; values of SEC and SEP were usually about 0.3 and 0.4 respectively. The use of regain instead of moisture content increases the standard errors slightly and in practice the instruments are calibrated to give moisture content, which is then converted to regain in the external computer.

Predictive equations for residual grease are quite different. In general it was found desirable to use 12 or even more variable terms. With residual grease being a minor component it seemed to be necessary to gather more information from the spectrum than is usual for "easy" variables such as moisture content.

The wavelengths chosen were selected purely on statistical grounds. No dire consequences arose from those procedures. The NIRA instruments ran 168 h/wk throughout the year, bias changes were not needed and recalibration was rare. The lack of ability to transfer equations was shown to arise from variations in the filter characteristics between the instruments [26] rather than from the use of long equations.

One unusual procedure used is the feedback of knowledge about the population. Williams [27] observed that when a calibration is based on a Gaussian distribution of samples with respect to concentration, then the results of the analysis of unknown samples would be regressed toward the population mean. The phenomenon makes one contribution to the skew often observed in the data because the high-concentration samples will be underpredicted and the low-concentration samples overpredicted. In practice skew from this source will be reasonably common for it is very difficult to attain the ideal rectangular distribution of concentration within the calibration set. This is certainly the case for the distribution of residual grease levels in scoured wools.

In the Gaussian situation Williams [27] claims that the NIR result will be related to the true result as follows:

$$N = \bar{P} + (M - \bar{P}) \times r$$

where N = NIR result

\bar{P} = population mean

M = manual (reference) result (i.e., "true" result)

r = correlation coefficient between the NIR result and the manual (reference) result for the calibration set

N is the NIR result arising from a calibration equation obtained in the usual way but the quantity of interest is M , the normally unknown manual (reference) result. The equation can be solved for M as follows:

$$M = \frac{1}{r}(\bar{P}(r - 1) + N)$$

Now

$$N = F_0 + \sum F_n \log \left(\frac{1}{R_n} \right)$$

is the normal predictive equation for an InfraAlyzer, obtained by the usual statistical methods. It follows that an improved predictive equation may be obtained by using the following parameters:

$$\text{New constant} = \bar{P} \left(\frac{r-1}{r} \right) + \frac{F_0}{r}$$

$$\text{New coefficients} = \frac{F_n}{r}$$

Both the calibration and the “unknown” sample sets must be representative of the same parent population for this step to be valid but this restriction is a normal requirement for NIR application anyway.

Examples of the efficacy of the correction are shown in Table 23.2 where the slope is the gradient of the line relating the residuals (reference result minus InfraAlyzer result) to the reference results. In all four cases a significant reduction has been produced. In two cases the corrected slope is not significantly different from zero.

The much greater number of data points (1050) obtained from the Model 6500 makes the use of BMDP inappropriate. All calibration work for the Model 6500 was carried out using the NSAS package from Foss-NIRSystems.

Comparisons were made [28] between the algorithms available within the NSAS package as well as with those provided by ISI-NIRS2 from Infrasoft International. These comparisons were made on data from calibration sets of 225 and 193 for residual grease and moisture content respectively. The corresponding validation set sizes were 123 and 109. It became clear that no one algorithm was consistently the best (lowest SEP); more commonly there was little to choose between them. The moisture content data of Table 23.3 are typical.

Similarly, for residual grease estimates, using “visible band” (400 to 1100 nm) data the results in Table 23.4 were obtained.

Clearly all of the algorithms are approximately equally efficient at extracting the relevant data from the spectra.

It can be noted too that no computational procedure is yet known to the authors that, when applied to wool data, outperforms MLR on $\log(1/R)$ from an InfraAlyzer 400.

The use of generously sized, carefully selected calibration sets together with high-grade reference data is mandatory. Elaborate computational procedures appear to add little.

The spectroscopic interpretations are not always straightforward. For example, consider the wavelengths selected by the NSAS package for moisture content as shown in Table 23.5.

TABLE 23.2
Efficacy of Skew Reduction Using William's Equation

Validation population	No. of samples	Slope ^a	CI
S14	124	0.219	±0.076
S14 corrected	124	0.130	±0.076
S18	193	0.110	±0.058
S18 corrected	193	0.007	±0.062
S32V	93	0.278	±0.105
S32V corrected	93	0.177	±0.107
S33V	150	0.199	±0.120
S33C corrected	150	0.089	±0.132

^a Slope: gradient of the regression line fitted to a plot of residual vs. reference values.

TABLE 23.3
SEP Values for Moisture Content

Software package	Equation	SEP
ISI	13 terms	0.27
NSAS	6 terms, 1st derivative	0.28
NSAS	3 terms, 2nd derivative	0.27
NSAS	3 terms, $\log(1/R)$	0.27

TABLE 23.4
SEP Values for Residual Grease

Software package	Equation	SEP
NSAS	6 terms, 7 factors (PLS)	0.081
NSAS	9 terms, 1st derivative	0.081
NSAS	9 terms, 2nd derivative	0.083

TABLE 23.5
Wavelengths Used for Moisture Content (NSAS)

Equation	Wavelengths
6 terms, 1st derivative	1460, 1520, 1760, 1940, 2020, 2100
3 terms, 2nd derivative	1480, 1520, 1820
3 terms, $\log(1/R)$	1740, 2020, 2180

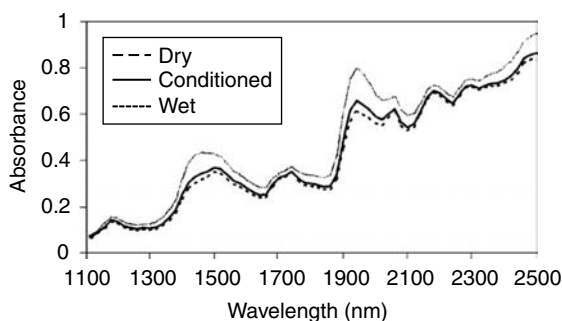


FIGURE 23.4 The spectra of one wool at three moisture-content levels.

The spectra of one wool at three moisture content levels are shown in Figure 23.4. Wavelengths clearly associated with water-related absorption have been chosen for each of the three-term equations of Table 23.5 with the exception of the 1740 nm choice in the $\log(1/R)$ set. That wavelength may be regarded as a reference wavelength, a common component of $\log(1/R)$ equations.

The case of the first derivative equation is less easy to explain yet it performs well. Moreover, such a situation prevails strongly for all residual grease equations. The spectral changes associated with residual grease levels are very subtle indeed.

Most, maybe all, workers derive their predictive equations using statistical approaches. Many, but not all, relate the results to spectroscopy afterward. Spectroscopy appears never to be used as the basis of selecting a wavelength in the first place. While this situation is less than fully satisfactory it can be noted there have been no unfortunate consequences observed by the authors, despite several years of industrial use of NIR equipment in several wool-scouring plants.

Calibration for “as-is” color measurements is best done by treating the instrument as a colorimeter rather than simply treating the color values as though they were constituents [29]. If this is impractical owing to software or other limitations then the use of transformed data will be helpful because colorimetric data such as tristimulus values X , Y , Z , are simply weighted sums of reflectances.

$$X = \sum_{\lambda} R_{\lambda} x_{\lambda} S_{\lambda} \Delta\lambda, \quad Y = \sum_{\lambda} R_{\lambda} y_{\lambda} S_{\lambda} \Delta\lambda, \quad Z = \sum_{\lambda} R_{\lambda} z_{\lambda} S_{\lambda} \Delta\lambda$$

Where R_{λ} is the reflectance of the object; $S_{\lambda} \Delta\lambda$ the spectral distribution of the flux irradiating the object; and x_{λ} , y_{λ} , z_{λ} , the CIE color-matching functions [30].

One useful transformation is to the variables R_x , R_y , and R_z using the following equations:

$$R_x = \frac{1}{78.2} \left(X - \frac{19.8Z}{118.1} \right)$$

$$R_y = \frac{Y}{100}$$

$$R_z = \frac{Z}{118.1}$$

R_x represents the amber component of X on a 0 to 1 scale; R_y and R_z are the corresponding Y and Z functions. It must be noted that the numerical constants in these equations relate only to Illuminant C data [30].

Calibration can then proceed using $\log(1/R_x)$, $\log(1/R_y)$, and $\log(1/R_z)$ as the reference data.

Other transformations that may be considered are those to the color-descriptive variables $L^*a^*b^*$.

$$L^* = 116 \left(\frac{Y}{Y_n} \right)^{1/3}$$

$$a^* = 500 \left(\left(\frac{X}{X_n} \right)^{1/3} - \left(\frac{Y}{Y_n} \right)^{1/3} \right)$$

$$b^* = 200 \left(\left(\frac{Y}{Y_n} \right)^{1/3} - \left(\frac{Z}{Z_n} \right)^{1/3} \right)$$

For Illuminant C-based data and the 1931 standard colorimetric observer with 10 nm summation $X_n = 98.041$, $Y_n = 100$, and $Z_n = 118.103$.

An example of the effects of these transformations is shown in Table 23.6. Note that all of the data have been retransformed to tristimulus values before calculation of the SEP values. All of the quoted NIR data have been obtained from a Foss-NIRSystems Model 6500.

The direct colorimetric approach favored for “as-is” color measurements is inappropriate for base color measurements unless the vigorous cleaning step of the standard method [21] is employed for the NIR test specimen. Calibration by correlation is employed. Examples of the SEP values obtained are shown in Table 23.7 together with the corresponding SE values for the reference method.

The morphological characteristics, fiber diameter and the degree of medullation, are sometimes requested. While an estimate of diameter may be obtained from Vis-NIR spectrophotometers, the data quality is quite inferior to that of the reference method and it is to be regarded as a guide only.

Conventional correlation approaches are used for each of them.

TABLE 23.6
“As-is” Color: SEP Values

Calibration	Y	Z
NSAS on Y, Z	0.97	1.21
NSAS on $\log(1/R_y)$, $\log(1/R_z)$	0.87	1.19
NSAS on L^* , b^*	0.89	1.12
“Colorimetric”	0.73	0.89

Number of samples: calibration 210, validation 110.
Range of values (validation set): Y 51–68, Z 41–63.
Reference measurements: Hunterlab ColorQuest. “Colorimetric” calibration: based on actual reflectances after correction. (M. J. Hammersley and P. E. Townsend, *Proc. 6th Intl Conf. on NIR Spectroscopy*, Lorne, Vic., Australia, 465 (1994).)

TABLE 23.7
Base Color: SEP Values

	SEP	SE (reference)
Tristimulus X	0.84	0.91
Tristimulus Y	0.86	0.92
Tristimulus Z	1.29	1.28

The SEP data refer to a Foss-NIRSystems Model 6500.

23.6 EXTENSIONS

The success of the WRONZ scoured-wool application has provoked a reexamination of applications previously discarded. Baxter and Wear [31], extending work by Nissen-Wooller and Marler [32], successfully argued for the incorporation of the NIR method as an acceptable alternative to soxhlet extraction with ethanol in the IWTO yield test method [1].

WRONZ has extended its calibration services to cover tops — a partly processed form of clean wool being essentially scoured wool plus lubricant. Greasy wool applications are also at an advanced stage of development.

Another step in the evolution of the scoured wool work would be to certify on the basis of in-scour NIR measurements rather than on soxhlet extractions at another locality some time later. The statistical argument is compelling. A commercial scourment may consist of 50 bales and each bale is sampled and tested by NIR. The certified soxhlet result is based on a small subsample of the pooled primary samples and so the certificate for perhaps 20 tons of wool is generated from two or three 10 g soxhlet test specimens while the scour has generated some 40 or 50 NIR results, one for each bale. The argument will be even more compelling after the widespread adoption of NiraSpec instruments generating their data on complete core samples.

While a change from certification by a separate organization to certification on the basis of in-scour measurements carries obvious risks, it is clear that there are no technical reasons for not changing.

Materials such as yarns, fabrics and carpets offer many possibilities for valuable applications of NIR methods. Wool is frequently blended with other fibers and control of the proportions may be

crucial. Enthusiasm for these needs to be tempered by the knowledge so well expressed by Davies [33], who has spoken of fully validated methods taking years rather than months to develop. Even so, the high stability of modern instruments means that recalibration can be infrequent and hence the rewards appreciable.

REFERENCES

1. IWTO Specifications of Test Methods, IWTO-19-85 (E), 1985.
2. H. R. Sabbagh and S. A. Larsen, *Proc. West. Am. Soc. Anim. Sci.*, 29: 108 (1978).
3. H. R. Sabbagh, M.Sc. Thesis, Univ. Wyoming, 1978.
4. R. F. Scott and E. M. Roberts, *Wool Tech. Sheep Breed.*, 26: 27 (1978).
5. R. F. Scott, E. M. Roberts, and M. J. Keogh, *Anim. Prod. Aust.*, 14: 515 (1982).
6. IWTO Specifications of Test Methods, IWTO-28-82 (E), 1982.
7. S. A. Larsen and J. L. Kinnison, *Text. Res. J.*, 52: 25 (1982).
8. S. L. Ranford, M. J. Hammersley, and V. C. Acker, *Wool Tech. Sheep Breed.*, 34: 147 (1987).
9. J. Lappage and J. Bedford, WRONZ Report No. R107, 1983.
10. S. L. Ranford, M. J. Hammersley, V. C. Acker, and F. L. Glassey, *Proc. 8th Int. Wool Text. Res. Conf.*, WRONZ, Christchurch, New Zealand, 1990.
11. T. Slack-Smith, D. Fong, and S. A. S. Douglas, *J. Text Inst.*, 70: 33 (1979).
12. J. P. Connell and G. H. Brown, *J. Text Inst.*, 69: 357 (1978).
13. J. P. Connell and K. H. Norris, *Text. Res. J.*, 50: 371 (1980).
14. G. J. Higgerson, J. W. Marler, and J. P. Connell, *J. Text Inst.*, 76: 133 (1985).
15. J. P. Connell and K. H. Norris, *Text. Res. J.*, 51: 339 (1981).
16. J. P. Connell, *Text. Res. J.*, 53: 651 (1983).
17. W. B. van Pelt, M. J. Hammersley, and G. Henderson, WRONZ Commun. No. 76 (1982).
18. M. J. Hammersley, *Proc. IWTO Tech. Cttee.*, Brussels, Belgium, Report No. 23 (January 1983).
19. M. J. Hammersley, WRONZ Communication No. C118 (1991).
20. P. E. Townsend and M. J. Hammersley, WRONZ Report No. R185 (1991).
21. IWTO Specifications of Test Methods, IWTO-56-99, 1999.
22. IWTO Specifications of Test Methods, IWTO-28-93 (E), 1993.
23. S. L. Ranford, M. J. Hammersley, and V. C. Patrick, *Proc. Int. Wool Text. Res. Conf.*, Tokyo, II, 167–175 (1985).
24. G. M. Furnival and R. W. Wilson, *Technometrics*, 16: 499 (1974).
25. W. J. Dixon (ed.), BMDP: Biomedical Computer Programs, University of California Press, Los Angeles, (1977).
26. M. J. Hammersley, S. L. Ranford, and P. E. Townsend, *Proc. Intl Wool Text. Res. Conf.*, Christchurch, II, 218 (1990).
27. P. C. Williams, *NIR 84, Proc. Int. Symp. NIRS*, Royal Australian Chemical Institute, Cereal Chemistry Div., Melbourne, 156 (1984).
28. M. J. Hammersley, P. E. Townsend, G. F. Grayston, and S. L. Ranford, *Text. Res. J.*, 65: 241 (1995).
29. M. J. Hammersley and P. E. Townsend, *Proc. 6th Intl Conf. on NIR Spectroscopy*, Lorne, Vic., Australia, 465 (1994).
30. CIE Colorimetry, 2nd ed. Vienna: Commission Internationale de L'Eclairage. CEI 15.2 (1986).
31. B. P. Baxter and J. Wear, *Proc. IWTO Tech. Cttee.*, Nice, France (December 1995).
32. B. F. Nissen-Wooler and J. W. Marler, *Proc. IWTO Tech. Cttee.*, Report No. 14, Punta del Este, Uruguay (April 1992).
33. A. M. C. Davies, *Eur. Spectrosc. News*, 73: 10 (1987).

24 FT/IR vs. NIR: A Study with Lignocellulose

Donald A. Burns and Tor P. Schultz

CONTENTS

24.1 FT/IR Analyses	479
24.2 NIR Analysis	481
24.3 Comparison	483
24.4 Discrimination	484
24.5 The Future	484
References	484

This study was prompted by a report on the rapid determination of lignocellulose by diffuse reflectance Fourier transform infrared reflectance (FT/IR) spectroscopy [1]. Although the results of that study were good, the claims by the promoters of near-infrared (NIR) instrumentation suggested that the latter might produce equally good results in less time and with less sample preparation.

Accordingly, samples that had been analyzed in the mid-IR with a Nicolet 20DX FT/IR instrument were reanalyzed (by DAB) in the NIR region with a Bran + Luebbe InfraAlyzer. The FT/IR analyses were performed by two different methods: (a) with a Harrick DRIFT attachment, and (b) via KBr pellet (with and without derivatization of the absorbance signal). This chapter compares the results obtained from the InfraAlyzer 500 with those produced on the Nicolet 20DX. Although both studies dealt with three analytes (lignin, cellulose, and hemicellulose) in two woods (pine, a conifer or softwood, and sweetgum, a hardwood), we limit this report to lignin only. Readers who require additional details are directed to the referenced literature.

24.1 FT/IR ANALYSES

Figure 24.1 is a tracing of the absorbance (and its first derivative) of sweetgum over the range 2000 to 800 wavenumbers (5.0 to 12.5 μm). The regression line obtained from the derivative treatment is shown in Figure 24.2 together with the individual data points from the two methods of analysis. The equations describing the line are as follows:

$$\begin{aligned} \text{A: } & -9.59 + 4(\text{Abs}_{[1472]}/\text{Abs}_{[1365]}) + 3.96(\text{Abs}_{[1602]}/\text{Abs}_{[1365]}) \\ & + 15.6(\text{Abs}_{[1311]}/\text{Abs}_{[1365]}), \quad R^2 = .962 \\ \text{B: } & -2 + 8.2(\text{Abs}_{[1602]}/\text{Abs}_{[1275]}) + 4.55(\text{Abs}_{[1587]}/\text{Abs}_{[1275]}) \\ & + 8.75(\text{Abs}_{[1016]}/\text{Abs}_{[1275]}), \quad R^2 = .968 \end{aligned}$$

The individual terms are ratios of absorbances at the wavenumbers specified within the brackets, and the correlation coefficient is given for each method.

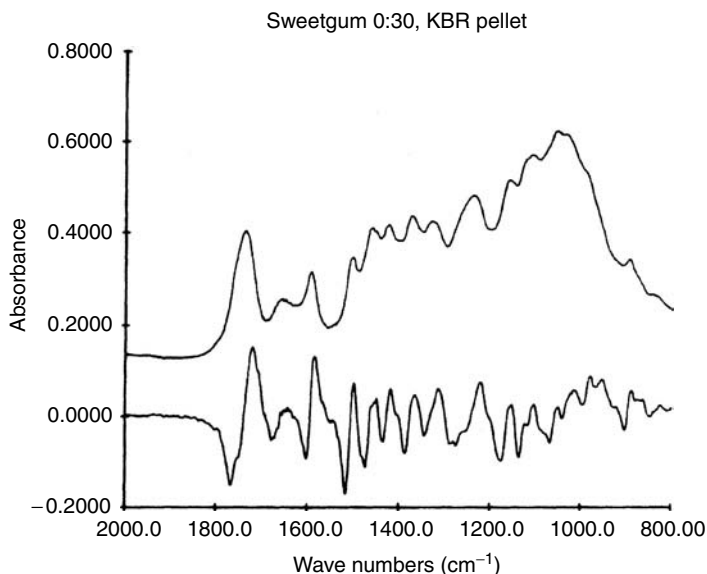


FIGURE 24.1 FT/IR spectrum of a typical sweetgum: absorbance and its first derivative.

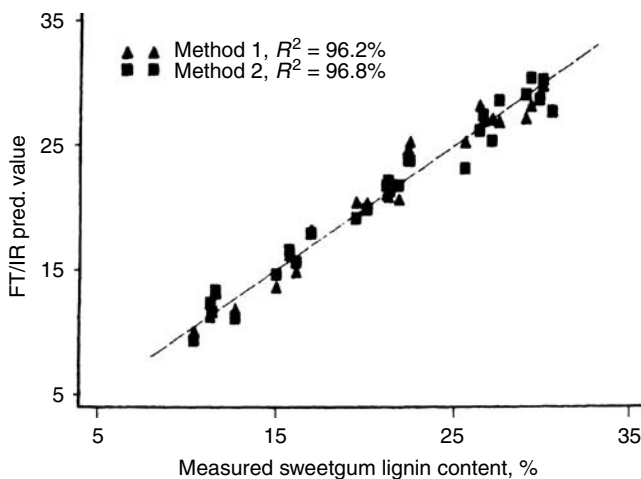


FIGURE 24.2 Lignin in sweetgum: predicted vs. actual.

Lignin in pine is a bit better. The regression is shown in Figure 24.3, and the corresponding equations are as follows:

$$\text{A: } 14 + 22.3(\text{Abs}_{[1212]}/\text{Abs}_{[1418]}) + 18.2(\text{Abs}_{[1365]}/\text{Abs}_{[1418]}), \quad R^2 = .977$$

$$\text{B: } 27.4 + 42.2(\text{Abs}_{[1212]}/\text{Abs}_{[1518]}), \quad R^2 = .981$$

Whereas the determination of lignin in sweetgum required equations with three terms (+ offset), its determination in pine required fewer terms for both methods. Despite this, the statistics improved.

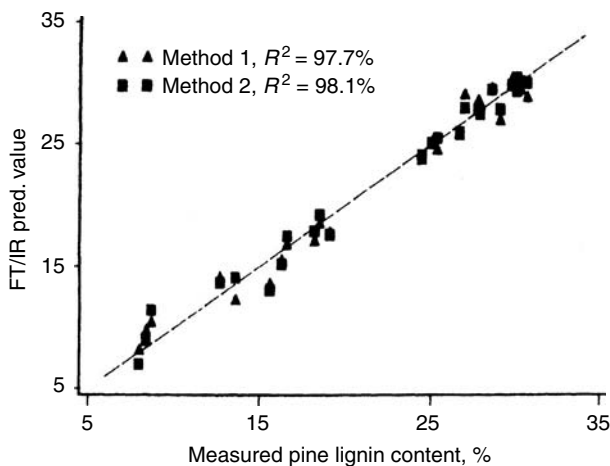


FIGURE 24.3 Lignin in pine: predicted vs. actual.

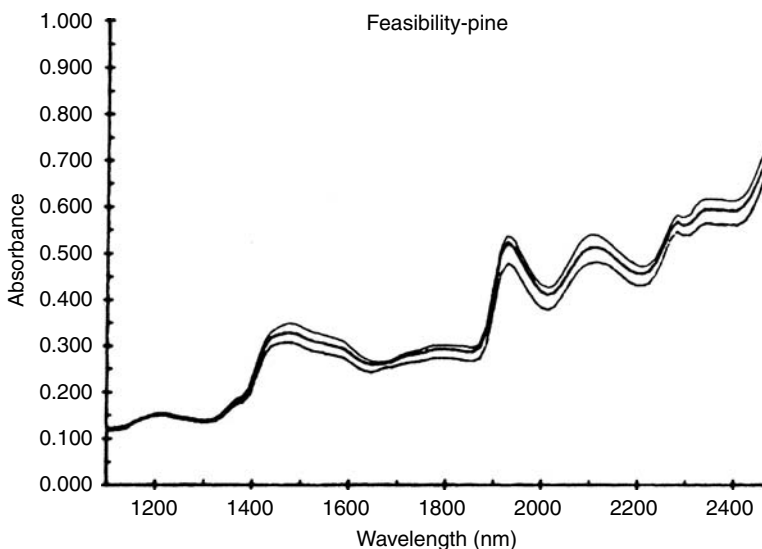


FIGURE 24.4 NIR spectra of pine samples containing low-, medium-, and high-levels of lignin.

24.2 NIR ANALYSIS

Sample preparation for NIR has been addressed elsewhere and will not be repeated here. Suffice it to say that an unmeasured sample is poured into a cup, the excess is smoothed off, the back is attached, the cup is placed into the drawer, and closing the drawer correctly positions the sample inside the instrument.

The reader should bear in mind that, since this study was done, several vendors have introduced FT/NIR instruments. If the initial effort had been done with one of these instruments, results would have been comparable, albeit likely very faster.

To see if NIR had a chance of competing with FT/IR, the initial study was one of feasibility. From the group of 33 pine samples we removed three that, from the chemical analysis, represented a high-, medium-, and low-level of lignin. The tracing in Figure 24.4 revealed that it had structure in the 1100 to 2500 nm region, and the visual differences among the three levels was encouraging.

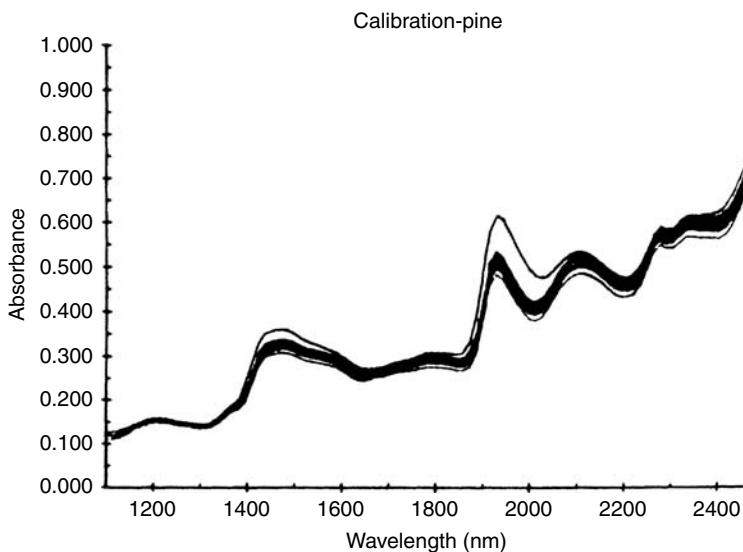


FIGURE 24.5 NIR spectra of 33 pine samples (including outlier).

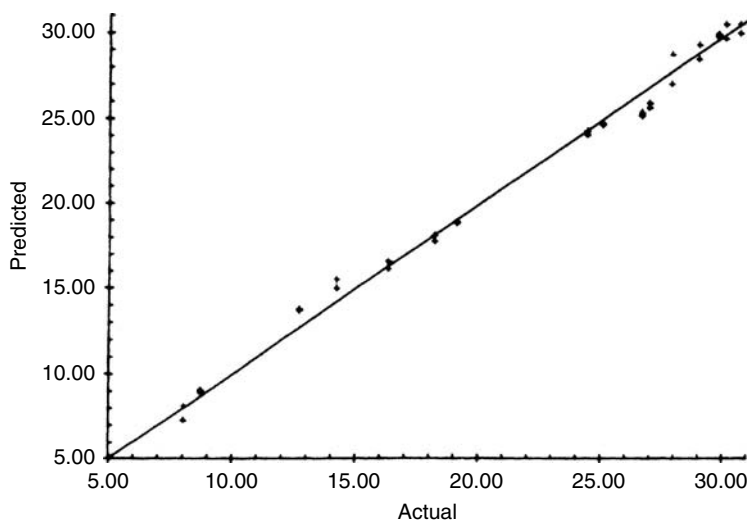


FIGURE 24.6 Lignin in pine: predicted vs. actual (calibration set).

Consequently, the entire group was scanned, producing the family of tracings in Figure 24.5. One sample was clearly different from all the rest. This was confirmed by a statistical analysis that identified it as an outlier, and it was removed from the set. The resulting regression line is shown in Figure 24.6, and its equation is as follows:

$$\% \text{ lignin} = 68.4 + 1334(\text{Abs}_{[2262]} - 987\text{Abs}_{[1828]}) + 639(\text{Abs}_{[2010]} - 1210\text{Abs}_{[2430]})$$

The correlation coefficient was .996 and the F value (a measure of robustness of the equation) was 2764. See the chapter dealing with statistics for an appreciation of this relatively high F value. Over the range 8 to 31% lignin, the standard error of estimate (SEE) was 0.41.

A second set of pine samples, not used for the calibration, was used as a confirmation set to test the validity of the equation. Figure 24.7 shows that the distribution of data points around its regression line is nearly as good as the calibration line.

The analysis for lignin in sweetgum produced similar results, but with slightly poorer statistics: three wavelengths were required (1478, 1492, and 1898 nm) to produce an equation with a correlation coefficient of .98, an F value of 490, and an SEE of 0.89 over the range 10 to 30% lignin.

24.3 COMPARISON

The two instrumental techniques are compared in Table 24.1 on the basis of their calibration statistics and the standard error of performance (SEP) with a confirmation set. NIR would appear to have an edge over FT/IR on all counts: correlation is better, equations are more robust (higher F values), and there are lower errors. If one adds to this comparison two time factors, NIR emerges as a clear winner, as is shown in Table 24.1.

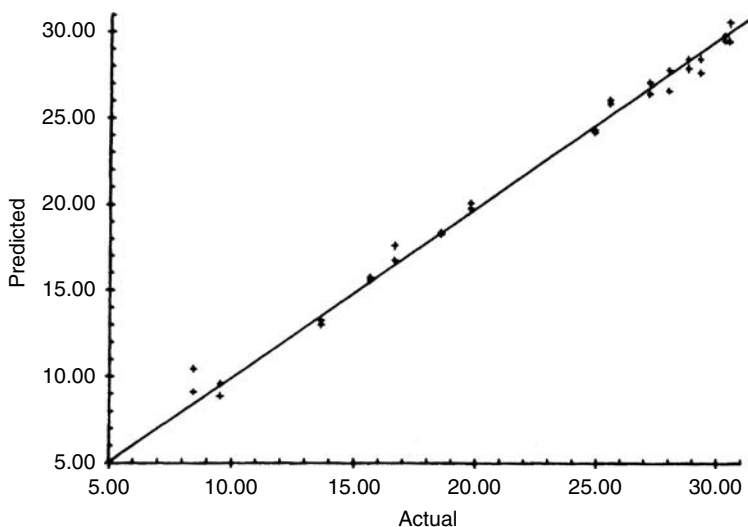


FIGURE 24.7 Lignin in pine: predicted vs. actual (confirmation set).

TABLE 24.1
Comparison of NIR and FT/IR for Lignin Analysis

Sample	Calibration					SEP (confirmation)			
	Corr coef.		F value			NIR		FT/IR ^a	
	NIR	FT/IR	NIR	FT/IR		Eq	Full	(A)	(B)
		(avg)		(A)	(B)				
Pine	0.998	0.979	2764	459	1152	0.55	0.64	2.05	1.43
Sweetgum	0.986	0.965	490	194	230	0.57	1.29	1.79	1.23

^a = 5 samples/group.

Eq = 8 samples, equivalent to FT/IR set.

Full = about 30 samples.

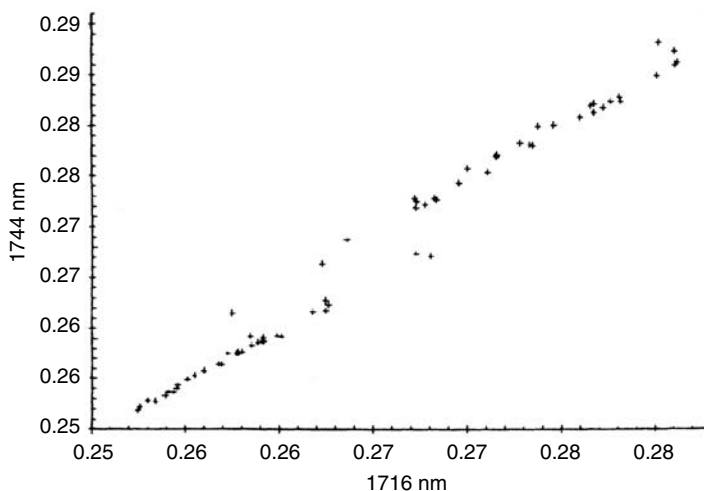


FIGURE 24.8 Pine and sweetgum grouping from discriminant plot.

We have reported [2] an extension of this study, which includes cellulose and hemicellulose in addition to lignin. NIR has less of an advantage when the matrix is pine and is somewhat inferior to FT/IR when the matrix is sweetgum.

24.4 DISCRIMINATION

Experts in this field could probably look at samples of pine and sweetgum and tell the difference instantly. But can the instrument tell the difference? Using a discriminant analysis program based on Mahalanobis distances, it was determined that three wavelengths are necessary and sufficient to distinguish the two groups. Figure 24.8 shows the distribution of all samples used in this study, plotted at the two most important wavelengths (1716 and 1744 nm). The two elongated ellipses neither touch nor overlap, although the addition of the third wavelength provides even more separation. Additional details on discriminant analysis are found in the chapter by Howard Mark.

24.5 THE FUTURE

Every year analytical instrumentation gets better, and this includes both FT/IR and NIR hardware. As already mentioned, we now have FT/NIR instruments to bring the advantages of FT down into the so-called NIR region. And “near” IR, which used to be defined as covering the range of 1100 to 2500 nm, now seems to include that space between the visible and NIR. It’s been dubbed the “far-visible” and the “near-near-IR” for lack of a more descriptive handle. Add to this the obvious advantages of fiberoptic probes for remote sensing, and we are destined to see instruments that give us “instant” answers — both the yes/no qualitative kind, and the very desirable quantitative variety. Finally, the chemometric data treatments (which many say got their start in the early NIR instruments, despite its not being so identified) get better along with the hardware. So it should not surprise anyone that the chemist of the future may get answers as fast as Robbie the Robot analyzed the corn squeezings in that 1956 sci-fi movie *Forbidden Planet*.

REFERENCES

1. T. P. Schultz, M. Templeton, and G. D. McGinnis, *Anal. Chem.*, **57**: 2867 (1985).
2. T. P. Schultz and D. A. Burns, *Tappi J.*, **73**: 209 (1990).

25 NIR Analysis of Textiles

Subhas Ghosh and James Rodgers

CONTENTS

25.1	Introduction	485
25.2	Natural Fibers	486
25.2.1	Amount of Reducing Sugar from Cotton Surface	486
25.2.2	Quantitative Analysis of Polyester/Cotton Blends	489
25.2.3	Degree of Mercerization	490
25.2.4	Cotton Fiber Maturity	494
25.2.5	Summary	496
25.3	Properties of Synthetic Fibers, Yarns, and Other Textile Products	497
25.3.1	Overview	497
25.3.2	Moisture Content in Nylon Fibers	498
25.3.3	Finish-on-Fiber Measurements	499
25.3.4	Heatset Temperature Measurements	501
25.3.4.1	Heatset Temperature Prediction	503
25.3.4.2	Process Control Measurements, Nylon Carpet Yarns	505
25.3.5	Textile Product Identification	506
25.3.5.1	Nylon and Heatsetting Type Identification	507
25.3.5.2	Polyester Fiber by Different Producers	507
25.3.5.3	Fiber/Polymer-Type Identification	507
25.3.5.4	Carpet Fiber Identification for Recycling	508
25.3.6	At-Line/On-Line Analyses	512
25.3.6.1	At-Line Measurement of Carpet Yarn Heatset Temperature	513
25.3.6.2	On-Line Measurement of PVA Size on Warp Yarns	515
25.3.6.3	At-Line/"Pseudo" On-Line Measurement of Tire and Carpet Yarn Moisture and Finish-on-Fiber	516
	References	519

25.1 INTRODUCTION

Near-infrared (NIR) spectroscopy has been found to be a useful technique to characterize raw materials and finished textile products, and NIR methods and techniques continue to find increasingly diverse and wide-ranging quantitative and qualitative applications in the textile industry. Quantitative analyses determine the amount (or quantity) of the property/species of interest in a substance or material. Qualitative analyses can be used to either identify a specific species or substance present in a material (i.e., coating on a fiber), the type of material itself (i.e., cotton, nylon, or polyester), or the quality of the material. NIR quantitative and qualitative methods allow the user to rapidly, accurately, and precisely monitor key chemical, physical, and morphological properties of textile fibers, yarns, fabrics, and chemical textile auxiliaries. Chemical properties are specific chemical species or groups present in the material (i.e., CH, OH, NH) that result in NIR spectral absorbencies at distinctive

wavelength(s). Physical properties are nonchemical molecular or physical properties (i.e., turbidity, light scattering) that result in changes or movement in NIR spectral absorbencies. Morphological changes are changes in the material's molecular structure or fiber's morphology (i.e., hydrogen bonding, crystalline state, and structure) that result in changes in NIR spectral absorbencies.

Near-infrared analyses are useful because a sample may be rapidly, accurately, and precisely tested without destroying its integrity or sample matrix. Because textile test specimens are normally in solid form, diffuse reflectance spectroscopy provides a special advantage by eliminating extensive sample preparation for testing textile materials. Additional productivity advantages are: simplicity and ease of operation, minimum training requirements, and multiple property/species analyses. A high level of accuracy and precision are achieved through the use of advanced chemometric software/modeling, stable instrumental platforms, and flexible instrumental settings (scan frequency, scan averaging, etc.).

This chapter will be presented in two sections — one for natural fibers (cotton, wool, blends, etc.) and one for synthetic fibers (nylon, polyester, blends, etc.). Each section will review many of the more common NIR analyses, methods, and techniques for textile products. Both laboratory and at-line/on-line NIR analyses will be reviewed. In addition, a more detailed description will be provided in each section for a few of the more common NIR textile techniques. Even though most textile readers may have prior knowledge of some of these detailed applications, it cannot be assumed that all readers are familiar with all of them. Thus, background and a brief preparatory description of the textile process and materials prior to each detailed application will be given, as well as specific references for those who wish to study the subject in greater detail.

25.2 NATURAL FIBERS

MRS has been found to be a useful technique to characterize raw materials and finished textile products. Because textile test specimens are in solid form, diffuse reflectance spectroscopy provides a special advantage by eliminating extensive sample preparation for testing textile materials such as fiber, yarn, and fabric. NIR analysis is useful because a sample may be rapidly tested without destroying its integrity.

Perhaps, some preparatory description of the textile process and materials prior to each application will be useful. Even though most textile readers are expected to have a prior knowledge of some of these applications, it cannot be assumed that they are familiar with all of them. Therefore, essential background is provided for each application as well as some references for those who wish to study the subject in greater detail. In this part of the chapter, a discussion of the measurement of cotton textile and synthetic textile properties is presented.

25.2.1 AMOUNT OF REDUCING SUGAR FROM COTTON SURFACE

Native cotton contains approximately 95% cellulose and some other materials such as waxes, pectins, protein, organic acids, and sugars. Stickiness of cotton during processing is often attributed to the high sugar-content level that is in excess of 0.5%. The problem becomes evident when deposits are built up on rolls and other parts of the machine. This situation is further aggravated by unfavorable climatic conditions. Most common observations show deposits on the cards, draw frames, combers, and spinning frames. In addition, more frequent roller lapping has been observed. Well-known causes of stickiness are the white fly (*Benesia tabi*) and aphid (*Aphis gossypii*) secretions commonly referred to as *honeydew*. Stickiness of cotton is also attributed to other plant sugars and seed oils [1]. Sugars mainly found in cotton include glucose, mannose, fructose, and pentoses; these are collectively called *reducing sugars*. The amount of reducing sugars and other related substances present in cotton is usually measured by comparing the reducing ability of the hot water extraction of the cotton to that of a reducing substance such as glucose.

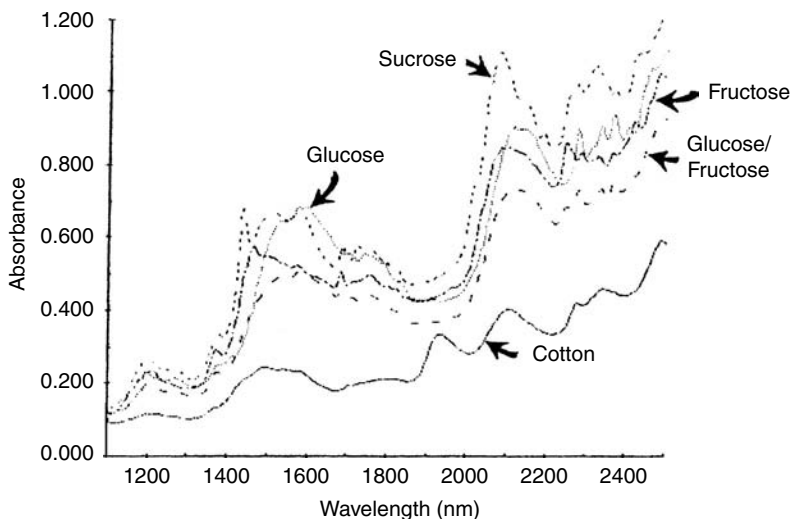


FIGURE 25.1 NIR spectra of cotton, glucose, fructose, sucrose, and a dry mixture of glucose/fructose.

Perkins' [1] procedure can be used as a reference method to calibrate the NIR instrument. In the Perkins' method, the aqueous sample extract of cotton, which mostly contains reducing substances, is reacted with an excess of potassium ferricyanide in the presence of sodium carbonate. The potassium ferricyanide oxidizes the reducing materials and in the process is reduced to potassium ferrocyanide. The amount of potassium ferrocyanide produced is determined by titrating with standard ceric sulfate in an acid solution using *o*-phenanthroline/ferrous sulfate complex (ferroin) as an indicator. The amount of ceric sulfate consumed is proportional to the amount of potassium ferrocyanide, which in turn corresponds to the amount of reducing materials present in original samples. Perkins' method is calibrated by using known amounts of glucose in standard solution.

Raw cotton contains both simple and complex carbohydrates. Simple carbohydrates include mono- and disaccharides. Collectively, water-soluble simple carbohydrates are classified as reducing sugars. Disaccharides are easily hydrolyzed to monosaccharides. In general, reducing sugars (monosaccharides) are often unbranched and regarded as straight-chain polyhydroxyaldehydes (aldoses) or ketones (ketoses); the former have the carbonyl group on C-1 (carbon atom 1) and the later on C-2. The remaining carbon atoms are hydroxylated. The mid-infrared examinations of reducing sugars show a broad band with shoulder in the region of 3.2 to 2.6 μm , which is due to OH groups that are hydrogen-bonded in various degrees. Because all reducing sugars have similar chemical structures, their NIR scans are expected to be quite similar, as shown in Figure 25.1.

Figure 25.1 was constructed by scanning cotton fiber and various sugars in powder form on a Technicon InfraAlyzer 500. Samples were scanned from 1100 to 2500 nm at 4-nm steps. It can be seen by the comparison of the spectra in Figure 25.1 that slight differences do exist, especially in the C—H combination region. Table 25.1 shows the principal bands in NIR spectra of raw cotton, sucrose, glucose, fructose, and a mixture (dry powder) of glucose and fructose (molar ratio 1:1). Band positions were obtained from an individual scan of cotton, glucose, fructose, and a mixture of glucose and fructose.

In Table 25.1, there are slight differences in the C—H combination and the C—H first overtone region of sucrose, glucose, and fructose. The assignment of absorption bands that are used for the generation of calibration equations can be made by finding the proper combination of the fundamental vibrations of functional groups associated with reducing sugars.

The coefficient of determination (R^2) was .98 with a standard error of prediction (SEP) of 0.02%. A 2-g cotton fiber in a solid sample cup was used for scanning. A four-wavelength regression model

TABLE 25.1
Principal Bands in NIR Spectra of Raw Cotton,
Glucose, Fructose, and a Mixture of Glucose and
Fructose^a

Cotton	Glucose	Fructose	Glucose/Fructose
2340	2340	2370	2370
2270	2255	2250	2260
	2150		
2100			2100
	2085	2080	
1920	1950	1945	1940
	1902		
1820			
	1730	1740	1740
1585	1590	1580	1580
1550	1520		1520
1490		1450	
1375	1385	1380	1380
1220	1195		1195

^a Wavelength expressed as λ_{max} (nm). Molecular ratio 1:1.

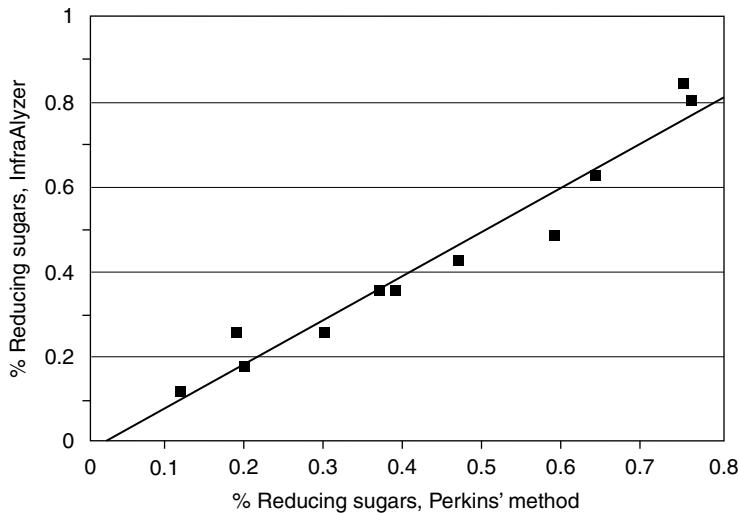


FIGURE 25.2 Relationship between the amount of reducing sugars using Perkins' method and as determined on the InfraAlyzer 500; $R^2 = .98$.

including 1445, 1722, 1905, and 2230 nm was generated from correlation of reflected $\log(1/R)$ data to the sugar values obtained from the titration method (Figure 25.2).

The absorbance bands occurring at 1445 and 1905 nm are assigned, respectively, to overcome and combination band of OH group stretching vibrations [2]. Usually, the free OH groups show greater absorbance due to the combination band around 1905 nm, while the hydrogen-bonded OH group exhibits maximum absorbance at longer wavelengths. In addition, the combination band is

more often utilized for the measurement of OH groups because its absorptivity is approximately twice as great as that of the first overtone band.

Giangiacoia et al. [4] used the NIR method successfully to analyze glucose, fructose, and sucrose from dry powdered samples. Reducing sugars present in aqueous solution were concentrated onto glass fiber by evaporation prior to analysis by the NIR instrument [5].

Because in aqueous solution there is more water–sugar than sugar–sugar interaction, Lanza and Li [6] were unable to use the NIR method to measure individual reducing sugars from fruit juices. However, by using slightly modified NIR methods that included separate calibration equations for each class of samples, these authors were able to analyze total sugars from a wide variety of fruit juices and wines.

Because cotton composition and physical characteristics vary widely from one growth area to another, it is necessary to collect the calibration sample set from a wide variety of cottons representing different growth regions. Cotton samples taken directly from the bale can be used for calibration and analysis. However, samples with a very high level of trash should be cleaned through a trash analyzer. A comparative measurement of reducing sugars in cotton by Near-infrared reflectance analysis (NIRA) and Perkins' manual method is shown in Figure 25.2.

The combination band at 2230 nm may be attributed to the CH₂ stretching and bending vibrations. This band position also corresponds to the combination band due to CO and OH groups stretching vibrations. Because the most reducing sugars show absence of a CO group absorption in the UV region, the combination band at 2230 nm may be entirely due to —CH₂ bending and stretching vibrations. The absorbance band at 1722 nm results from the first overtone of CH stretching vibrations. Apparently more than one assignment is possible to give the observed band because of the complex sugar scans in the NIR region.

25.2.2 QUANTITATIVE ANALYSIS OF POLYESTER/COTTON BLENDS

It is a common practice in textile manufacturing to blend fibers of two or more generic types to achieve desired fabric characteristics. For example, cotton is often blended with polyester terephthalate (PET) fibers to enhance durability and make the fabric easy to care for. The proportion of each type of fiber in the total blend must remain constant in the manufacturing process. Normally, polyester/cotton blend analysis is performed by dissolving cotton in 70% H₂SO₄ solution. This procedure requires 8 h before the final result is obtained. NIRA is used to perform a blend analysis within 2 min in textile laboratories.

Both cotton and polyester (polyethylene terephthalate) fiber show characteristic bands in the NIR region. Principal bands in NIR spectra of cotton and polyester are listed in Table 25.2. Band positions are calculated from the individual spectra of cotton and PET, as shown in Figure 25.3. Usually a regression model consisting of three or four wavelengths can accurately determine the composition of a blend.

Two methods of blending are employed in the textile industry: intimate and drawing blend. Intimate blending is performed at the very beginning of the process in the opening line, and draw blending is done at an intermediate processing stage in the drawing operation. The sample presentation differs slightly, depending on the type of blended stock, because the homogeneity of the blend depends on the method of blending. Intimate blend samples require two readings with different sample orientations with respect to the source (0 and 90° rotation); an average of the two readings should be used. Draw blend samples, however, require four determinations: after two determinations (at 0 and 90°) the specimen must be turned over and two more readings taken at different orientations.

A set of samples was analyzed using the NIR and a gravimetric method to determine the percent polyester in the blend, as shown in Table 25.3. Because the complex NIR spectra of the polyester/cotton is difficult to analyze, a *spectra reconstruction* [7] method can be used to generate separate spectra for polyester and a spectra of cotton for a spectrum of the polyester/cotton

TABLE 25.2
Principal Bands in NIR Spectra of Cotton and Polyester Terephthalate

Cotton, λ_{\max} (nm)		Polyester terephthalate, λ_{\max} (nm)		
1216	1776	1128	1828	2256
1270	1824	1172	1868	2336
1372	1930	1368	1908	2396
1444	2104	1412	1952	
1490	2276	1616	2088	
1550	2340	1660	2132	
1590		1716	2156	
1708		1800	2184	

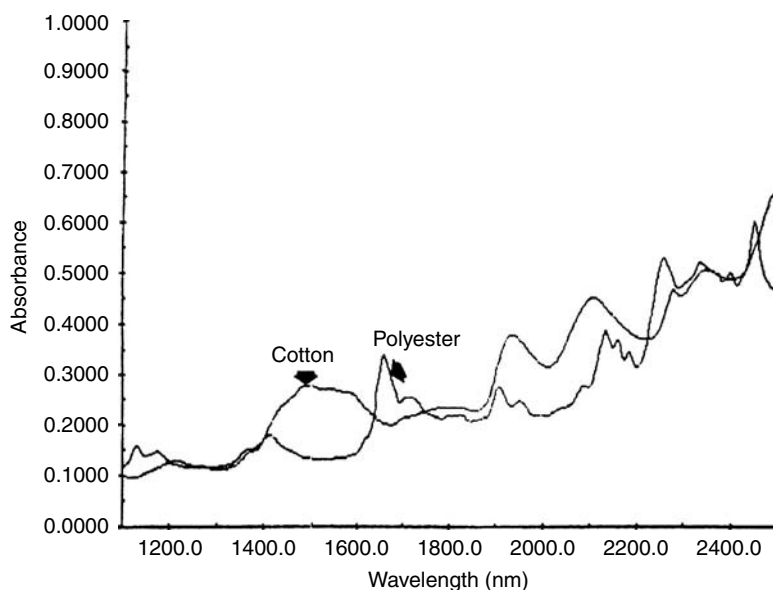


FIGURE 25.3 NIR spectra of polyester (PET) and cotton.

blend sample. An example of the spectral reconstruction is shown in Figure 25.4 that can be used to interpret the method.

25.2.3 DEGREE OF MERCERIZATION

In the mercerization process, cotton fabrics are usually treated with 20% NaOH solution under tension. Its purpose is to enhance the fabric's characteristics such as dye affinity, dimensional stability, tensile strength, and lustre [8,9]. The process must be controlled to ensure fabric quality. Dye shade variation is the most common quality problem related to mercerized fabrics. The conventional method for determining the degree of mercerization, barium activity number (BAN), is laborious and requires 6 h before a result is obtained, which makes the test unsuitable for process control. Determining the BAN has been the only accepted method for measuring the degree of mercerization. The BAN of mercerized fabric is determined by boiling the samples of mercerized and unmercerized fabrics in

TABLE 25.3
Blend Analysis

Percent PET H ₂ SO ₄ dissolution	Percent PET by NIRA
54.15	54.11
57.31	57.10
58.48	58.27
64.48	65.25
68.85	69.05
67.64	66.18
73.09	73.53
74.06	74.87

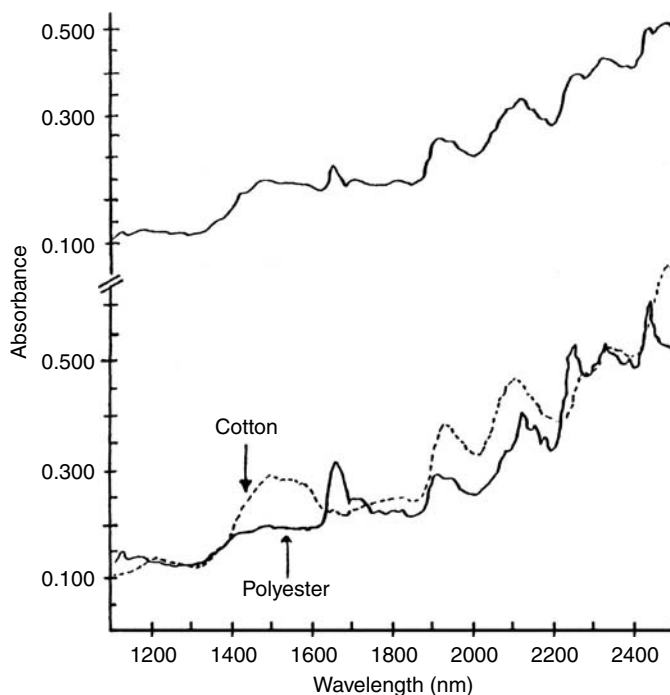


FIGURE 25.4 Spectral scan of 50/50 polyester/cotton blend (A) and spectral reconstruction of polyester/cotton blend (B).

a soap and soda ash solution. Both samples are washed until a neutral pH is obtained. Fabrics are treated with a barium hydroxide solution for a definite time period. The ratio of the amount of barium hydroxide absorbed by the mercerized specimen to that absorbed by the unmercerized specimen is determined by titrating against 0.1 HCl.

NIR spectra of unmercerized and mercerized cotton have been compared by Nikitin [10]. It has been stated that mercerization caused band shifts at specific wavelengths and the weakening of the intermolecular interaction of the cellulose structure along with an increase in the OH—O spacing. The band shifts, however, could also be attributed to a structural rearrangement of the cellulose macromolecules. Untreated native cotton contains mainly cellulose I structure; however, conversion of cellulose I to cellulose II occurs when the native cellulose is treated with NaOH. The conversion

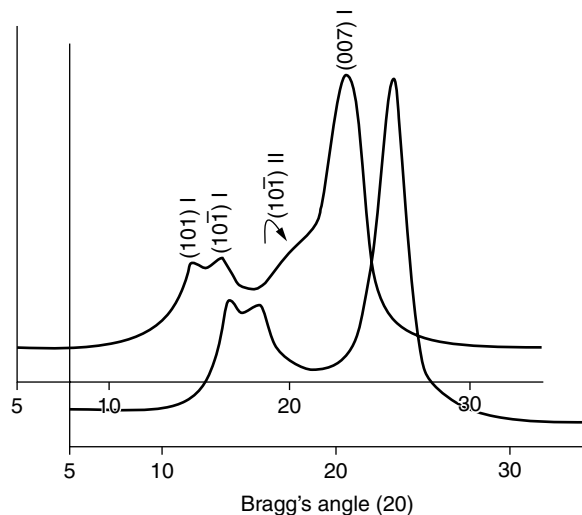


FIGURE 25.5 Comparison of equatorial scans of cellulose I and forming cellulose II.

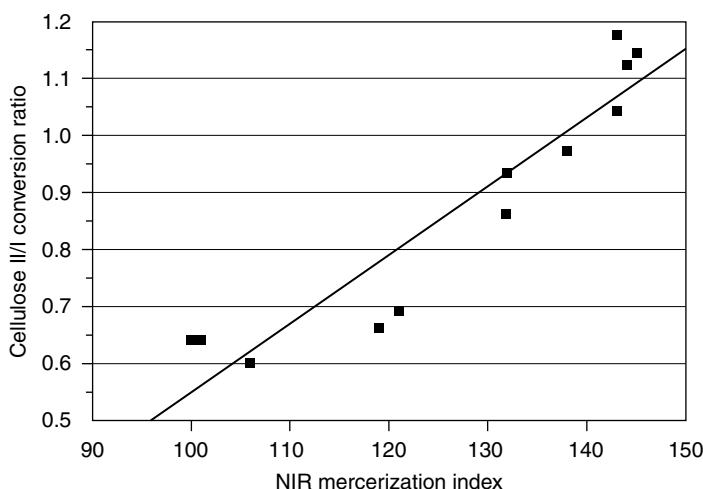


FIGURE 25.6 Relationship between the cellulose II/I conversion ratio and mercerization index.

is associated with crystal lattice transformation and a decrease in order due to the imperfection in the cellulose crystallites. Equatorial x-ray diffraction trace of the mercerized and unmercerized cotton, which was prepared using an Ni-filtered $\text{CuK}\alpha$ radiation, is shown in Figure 25.5. It is interesting to observe that a shoulder begins to appear at $2\theta = 20^\circ$, consisting of the [10] line of cellulose II in the mercerized sample. The degree of conversion of cellulose I to II, owing to the NaOH treatment, is illustrated in Figure 25.6. The conversion ratio was calculated from the diffractogram after correcting for the change in order using a procedure reported by Gjønnes and Norman [11].

A three-wavelength regression model can be used to determine the degree of mercerization of cotton fabric. There is an excellent agreement between the BAN and NIR mercerization index as illustrated in Figure 25.7.

NIR spectra of native and mercerized cotton are shown in Figure 25.8. The hydroxyl group overtone region near 1445 nm shows a more pronounced band in mercerized cotton and some of the

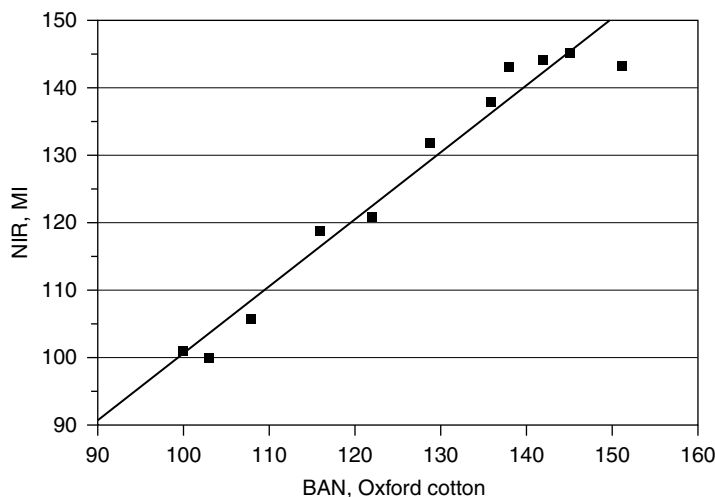


FIGURE 25.7 Relationship between BAN and mercerization index measured by NIRS.

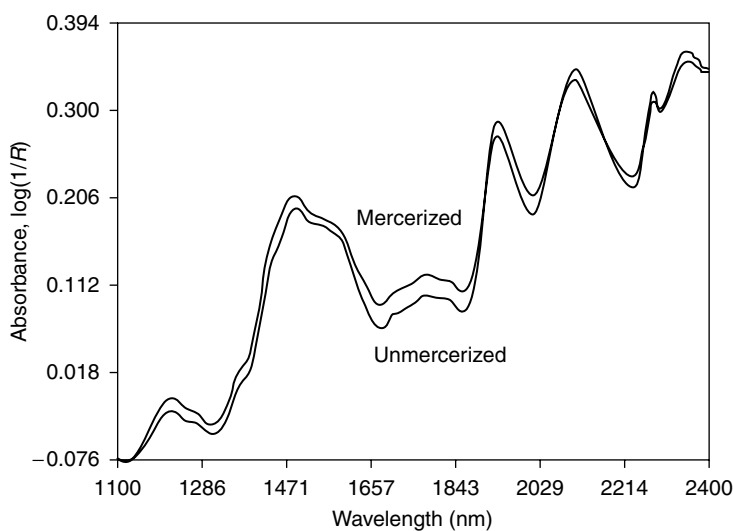
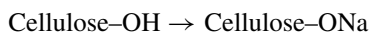


FIGURE 25.8 NIR spectra of native (low) and high mercerized cotton.

bands have disappeared because of the NaOH treatment by removing various noncellulosic materials from the cotton surface. The second-derivative spectra of the mercerized cotton (Figure 25.9) show a change in peak ratio in the hydroxyl overtone region, indicating a conformational change of the OH group. Moreover, there is a change in absorption at the 1820 nm band as the degree of mercerization increases.

It is assumed [10] that an alcoholate-type derivative is formed:



when the cotton fabric is treated with NaOH. After rinsing with weak acid and neutralizing the original structure, cellulose-OH is retained; however, the conformation of the macromolecule changes and thus the hydroxyl groups involved in the hydrogen bonding are affected.

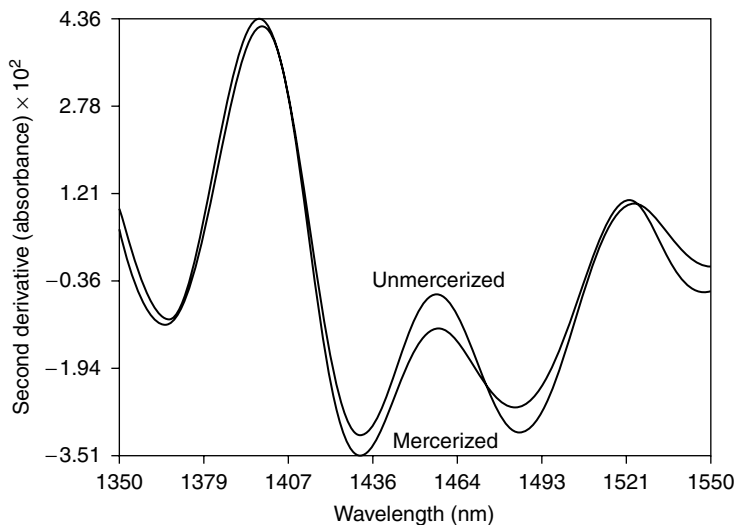


FIGURE 25.9 Second-derivative spectra of high- and low-mercerized cotton.

25.2.4 COTTON FIBER MATURITY

Cotton is by far the most important fiber in the apparel textile industry. Cotton fiber maturity is an important fiber property that influences end use properties of raw cotton. Fiber maturity does not indicate the length of the period of fiber growth. It means that the fiber wall has developed to an acceptable level of thickening. The cotton fiber initially grows out from the seed wall with just the primary (outside) wall being formed. Only after the fiber has grown to almost its full length does the secondary wall begin to form, starting on the outside wall and growing inward, causing the fiber to thicken. Immature cotton results when the normal wall thickening processes are interrupted because of insects, plant disease, and bad weather. The degree of maturity of cotton fiber must be determined before it is processed in the textile plant because immature cotton produces an unacceptable product.

A typical chemical composition of native cotton is given in Table 25.4 [12]. The main constituent is cellulose. Some other materials, such as waxes, pectin, sugars, organic acid, protein, and so forth, are also present. The relative amount of pectins and other noncellulosic substances reduce as the cotton matures [12].

Ward [13] pointed out that crystal growth occurs during the maturity of cotton fibers. Therefore, better order and greater crystallinity of the cellulose molecule is expected for the matured fibers.

Usually, two methods are used to determine cotton fiber maturity, for example, Causticaire and the Shirley Maturity Tester. These time-consuming methods also confuse fiber fineness with maturity. Therefore, they provide biased maturity values. A more fundamental method has been developed by Thibodeaux, USDA, where maturity of cotton is measured by taking the ratio of the fiber wall area to the area of a circle of the same perimeter. This is accomplished by using a microscope and an image analyzer. NIR is a better alternative in determining fiber maturity because it can be calibrated to measure maturity at different levels of fiber fineness and is fast in testing as a routine procedure. Two important criteria must be considered while selecting samples (a “trial set”) for calibration. Characteristics of cotton vary considerably from one growth area to another and from one generic variety to another. Therefore, a trial set should contain samples from different growth areas and varieties to obtain a robust calibration. Moreover, it is necessary to categorize samples in different groups of fiber fineness to use an appropriate bias for each group. A sample presentation does not require any special technique. A 2-g sample can be packed in a standard solid sample cup for presentation.

TABLE 25.4
Composition of Typical Mature
Cotton

	Percent
Cellulose	94.0
Protein ($N \times 6.25$)	1.3
Pectic substances	0.9
Ash	1.2
Wax	0.6
Malic, citric, and other organic acids	0.8
Total sugars	0.5
Pigment	Trace
Other	0.9

Source: H. Howell and J. Davis, *Textile Chem. Color*, 23: 69 (1991).

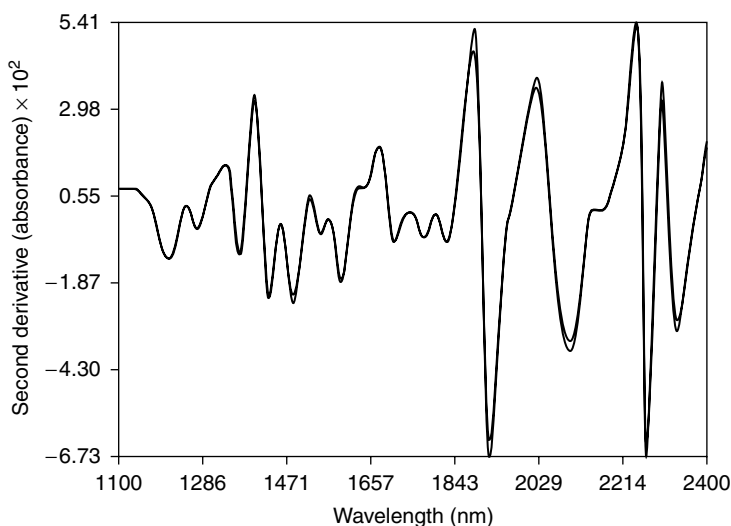


FIGURE 25.10 Derivative spectra of matured and immature cotton.

In Figure 25.10, second-derivative spectra of immature and matured cotton are illustrated. Absorbance change due to a change in maturity can be observed at various wavelengths. More than one model can be used to predict fiber maturity. However, a four-wavelength regression model was developed by correlating absorbance $\log(1/R)$ values to microscopic maturity values.

Fiber samples collected from the various textile plants were tested by the NIR and microscopic methods. The relationship between the two sets of data is shown in Figure 25.11, where agreement was excellent with $r^2 = .97$. NIR cotton maturity tests were performed on four varieties of cotton during their various stages of growth as illustrated in Table 25.5. This method of determining cotton maturity provides a technique for textile manufacturing to quickly detect cotton samples that could create quality problems in textile products.

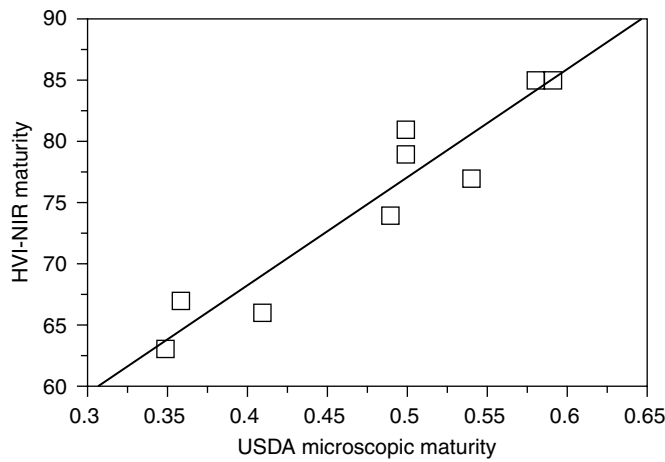


FIGURE 25.11 Relationship between the ITT NIR maturity and microscopic maturity by USDA.

TABLE 25.5
Cotton Fiber Growth Period and Maturity

Days after pink flower	NIR maturity index			
	STVL 213	STVL 825	DPL 50	DBS 119
25	50.2	42.7	47.4	49.5
30	63.1	59.7	57.7	63.7
35	64.0	60.9	61.6	66.0
40	68.9	68.8	69.1	69.4
45	69.9	75.9	72.0	68.6
50	75.5	79.2	76.7	75.0
55	75.7	78.7	77.5	77.5
<i>r</i>	.97	.98	.99	.95

25.2.5 SUMMARY

Near-infrared reflectance analysis is a useful technique for characterizing textile raw materials, fiber, yarns, and fabrics. It is a nondestructive quantitative analysis that is simple to use and allows rapid testing of the sample. Its ability to measure multiple components of the sample simultaneously and eliminate extensive sample preparation are major advantages of NIRA in the characterization of textile materials. Many innovative mathematical treatments, for example, discriminant analysis and spectral reconstruction, have been developed by instrument manufactures and software companies. These instruments not only aid in the quantitative analysis of the data but also allow morphological investigations of fibers and yarns and rapid, qualitative identification of specific sample sets.

Several NIRA applications were presented on the quantitative characterization of textile raw materials, fibers, yarns, and fabrics. NIRA has been used to measure the percent moisture and finish-on-fiber (FOF) of fibers and yarns, the amount of reducing sugars and maturity of cotton fibers, the degree of mercerization of cotton fabrics, the percent polyester in polyester/cotton blends, and the heat-set temperature of nylon carpet yarns. In each analysis, excellent agreement has been observed between the standard laboratory method and NIRA results. Various mathematical treatments of the data, for example, derivative math, enable the investigator to predict the probable type and extent of morphological changes occurring within the textile samples.

Discriminant analysis has been used to identify and classify similar but different fibers, polymers, and yarns based on the differences in their properties and morphology. It has been used to identify and classify different nylon polymer types (nylon66 and 6) and heat-setting methods (Suessen and Superba) by polymer type, heat-setting method, and simultaneously by polymer type and heat-setting method. Further, discriminant analysis has been used to identify and classify staple polyester fibers by fiber producer, by tenacity level, and simultaneously by producer and tenacity level.

As the advances in NIRA methodology and instrumentation continue, increased use of the technique for textile applications will also continue to grow. One area of high potential is that of online real-time analysis of textile samples in a production line. NIRA exemplified the use of sophisticated late twentieth century instruments in an established, relatively conservative industry. It does so by characterizing textile materials accurately, quickly, with simple and easy-to-use procedures.

25.3 PROPERTIES OF SYNTHETIC FIBERS, YARNS, AND OTHER TEXTILE PRODUCTS

25.3.1 OVERVIEW

Numerous quantitative and qualitative NIR techniques have been developed and reported for synthetic fibers, yarns, and other textile products. Many of these applications can be grouped into five major categories — heat/thermal history, “coatings,” fiber identification, moisture content, and at-line/on-line measurements. Several authors have used NIR techniques and methodology to monitor the heat history of synthetic fibers and yarns. Ghosh, Rodgers, Tincher, and coworkers [14–18] reported on NIR techniques to measure the heat history and heatset temperature of nylon carpet yarns. Ghosh and Tincher also studied the heat history and setting temperature of polyester (PET) textured yarns and carpet yarns [19–20].

NIR investigations and monitoring of the level of coating applied to the textile substrate (FOF, binder, size, lube, etc.) were performed by Drews, Lemere, Rodgers, Hill, and Harper and coworkers. Drews [21] examined by NIR the latex binders on PET nonwoven materials. Lemere [22] applied NIR techniques to monitor warp size add-on and composition for PET blends. Rodgers [17] reported on the NIR measurement of FOF applied to nylon yarns. Hill [23] measured the lube/coating level applied to sewing thread. Harper [24] studied the application of FOF to rayon yarns.

Fiber identification includes not only the specific type of polymer (nylon66, PET, etc.) and processing (i.e., type heatsetting) for a textile substrate, but also can include the quality and consistency of the material relative to standard production and previous historical samples. Howell and Davis [25] reported on the identification of dyed and greige fibers using NIR spectroscopy. Ghosh and Rodgers [26] used NIR Discriminant Analysis to differentiate between nylon66 and nylon6 carpet yarns that had been either Suessen or Superba heatset — joint identification of polymer type and heatset processing type. Ghosh [27] identified and classified staple PET by producer and tenacity level with the use of NIR Discriminant Analysis. Considerable research has been performed in the identification of the polymer type in textile materials — especially carpets — for recycling purposes. Costello and Knepper [28] discussed a new, customized NIR analyzer that quickly and accurately determined the carpet fiber polymer type in a manufacturing environment. Jasper and Kovacs [29] used neural networks and NIR spectroscopy to identify various synthetic and natural fibers from 17 fiber groups. Beck and coworkers [30] also developed a fiber identification library that differentiated between 18 fiber groups of synthetic and natural fibers (390 samples), using Principal Component Analysis (PCA). Rodgers [31] investigated the influences of carpet and instrumental parameters on carpet face fiber identification by NIR, using both static/laboratory and fiber-optic probe sampling systems.

One of the most thoroughly researched applications of NIR to textile substrates is the measurement of substrate moisture content. Most of the textile moisture applications are for natural fiber products — wool, cotton, and so forth. For synthetic textile products, Elliott and coworkers [32] studied the spectral locations and the spectral influence of water on synthetic fiber spectra. Rodgers [17,33] measured the moisture content of nylon fibers by NIR spectroscopy both in the laboratory and at-line/“pseudo” on-line.

One of the most recent applications of NIR methodology for textile substrates is in the area of at-line and on-line analyses. At-line analyses are those in which the NIR analyzer is located in the manufacturing area and the samples are brought to the analyzer for analysis by area operators rather than sending the sample to the lab. On-line analyses are those in which the NIR analyzer is placed in the manufacturing process and the NIR measurement, sample handling, and sample preparation are controlled automatically by a computer. Rodgers and Ghosh [34–37] have reported on the transference of the NIR measurement of nylon carpet yarn heatset temperature and heat history from the lab to at-line in the manufacturing area. Ghosh and coworkers [38,39] measured on-line the size content on PET blend warp yarns and the durable press content on PET blend fabrics. Jerry [40] studied the pigment concentration during colored fiber extrusion using on-line NIR spectroscopy. Rodgers [33] monitored the moisture content and surrogate FOF of textile bobbins during spinning using a portable NIR analyzer.

To explore how NIR methodology can be used for textile products and to demonstrate the versatility and robustness of the technology, a more detailed description is provided in Section 25.3.2 to Section 25.3.6 for a few of the more common NIR textile techniques. Example textile processes and measurements are discussed and reviewed for the areas of moisture content, FOF, heatset temperature, textile product identification, and at-line/on-line analyses.

25.3.2 MOISTURE CONTENT IN NYLON FIBERS

The moisture content in synthetic fibers, yarns, and fabrics is a critical variable that can have significant impact on physical properties, processing behavior, and manufacturing productivity. For hydrophilic materials, like nylon, the moisture content in the fiber, yarn, and fabric serves as a complementary variable to temperature and heating time [41,42]. It is known that temperature and heating time significantly influence the morphology of the synthetic fiber, which in turn impacts its dyeing and physical properties. For nylon fibers, it has been shown that an increase in the water content (i.e., increased relative humidity or increased moisture add-on in processing) of the fiber yields similar results as with an increase in temperature or heating time. Hence, it is of critical importance in the textile manufacturing process to monitor and control the moisture content of the nylon material.

There are several methods that can be used to measure the moisture content of nylon materials. These methods include such well-known techniques as titration, dielectric capacitance, resistance, heated balance, and microwave devices to name a few. One of the better-known standard methods for determining the total moisture content of textile fibers and yarns is the Karl Fischer Reagent (KFR) titration method. The KFR method is the quantitative reaction of water with the methanolic KFR solution of I_2 , O_2 , and pyridine (or pyridine substitute) [43]. The method consists of placing the yarn sample in a flask containing the KFR solution and then back-titrating the system with KFR solution to match a blank's (no yarn) color. The end point of the titration can be visually observed in the back titration by the color change from yellow to rust brown, or it can be determined coulometrically. Although the KFR method is very accurate, it is time consuming, requires careful laboratory and sampling techniques, and normally uses odorous chemicals. NIR reflectance analysis can be used to perform the analysis on nylon yarns in less than 5 min in a textile laboratory and avoids the aforementioned difficulties.

Water has very strong absorbance bands in the NIR spectral regions [44–46]. A laboratory investigation comparing the percent moisture on nylon66 spun fiber as measured by NIR to that

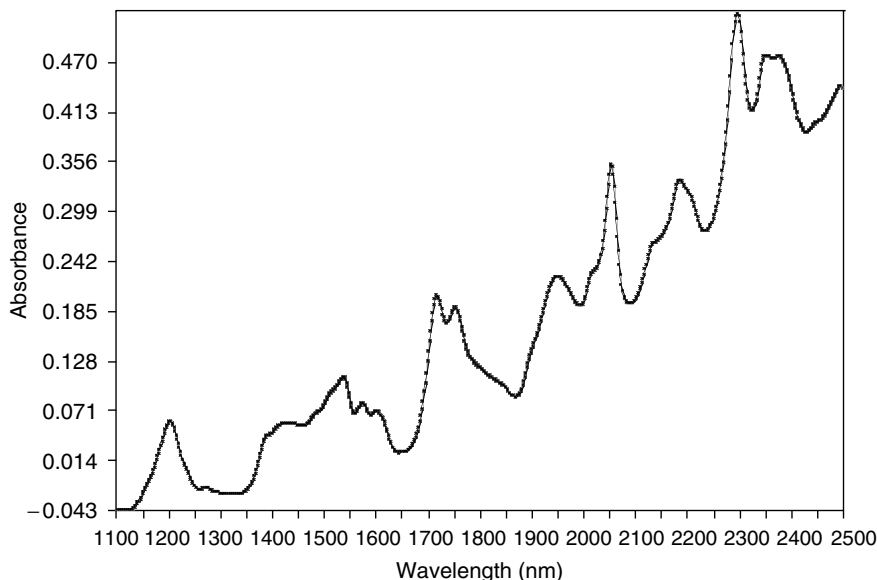


FIGURE 25.12 Percent moisture on nylon66 spun yarn by the NIR and KFR methods, three-wavelength model. (Redrawn from information in Rodgers, *Proc. 7th International NIRA Symposium*, Technicon Technical Center, Tarrytown, NY, 1984.)

measured by the KFR method is shown in Figure 25.12 [17]. The sample size was 3.5 ± 0.5 g and two packings were made for each sample. The calibration samples ranged from 0.60 to 10.87% moisture. From the calibration data, the regression models used in the investigation were a two-wavelength (1722, 2083 nm) and a three-wavelength (1445, 1905, 1940 nm) model. The SEP were excellent, being 0.5% H_2O for the three-wavelength model and 0.3%, for the two-filter model. The three-wavelength model was expected, as the 1445, 1905, and 1940 nm wavelengths have been previously assigned as OH group stretching vibrations. The precision and sensitivity of the two-wavelength model was unexpected. The 1722 nm wavelength has been assigned as the first overtone of CH stretching vibrations and the 2083 nm wavelength has been assigned as the $\text{C}=\text{O}/\text{N}-\text{H}$ stretch combination. Since neither wavelength corresponds to an expected OH band assignment, the two-wavelength model must be responding to a particular change in the nylon66 yarn rather than the moisture level alone. Since nylon66 is a hydrophilic material, it has a strong attraction for water and is capable of having hydrogen bonding between the water and nylon66 molecules. Thus, the hydrogen bonding forces and molecular structure of nylon66 are somewhat dependent on the percent moisture content. One possible explanation for the two-wavelength model is that changes in the percent moisture content of the fiber result in changes in the crystalline environment of the CH and amide species, and it is this change that the NIR method is detecting. An at-line/on-line measurement of fiber moisture will be presented in the At-Line/On-Line Analyses section.

25.3.3 FINISH-ON-FIBER MEASUREMENTS

It is a common practice in the manufacture of synthetic fibers to add a small amount of lubricating material to the fiber to assist its performance and runnability in downstream textile manufacturing processes [41,47–49]. The lubricating material is often referred to as “finish.” The three primary functions of the finish are lubrication of the fiber surface, static protection, and filament cohesion. The thin layer of finish is an interface between the fiber and other surfaces. The finish acts as a lubricant to reduce frictional wear and damage when the fiber passes over other contact surfaces,

such as guides, rolls, and needles. The finish provides static protection and filament adhesion to the fiber bundle so as to maintain the fiber bundle integrity and uniformity, properties that are critical to the success of the downstream textile manufacturing of the fiber bundle. The two basic methods for finish application to the fiber surface are the roll and metered finish methods. The more traditional roll application consists of a rotating roll applicator, partially submerged in the finish solution, taking the finish solution from the roll to the fiber surface. This application method is very simple and inexpensive but normally not uniform in its application of finish. The metered finish method consists of an accurately regulated flow, or “spray,” of finish through a metering pump directly onto the fiber surface, yielding a more uniform application of finish. It is also desirable to “spread” the fiber bundle prior to finish application, so individual filaments are coated as uniform as possible.

The amount of finish on the fiber surface, called “Finish-on-Fiber” or FOF, is of critical importance in the textile manufacturing process. Too little finish on the fiber will not give the fiber bundle acceptable lubrication, static protection, and filament cohesion for downstream processing, resulting in “flaring” and excessive mechanical wear on the fiber bundle that will cause interruptions and lost productivity, in downstream processing. The application of excessive finish to the fiber may cause the fiber bundle to adhere to contact surfaces, causing interruptions in downstream processing, and often results in deposits on contact surfaces that not only affect the downstream process efficiency but also lead to increased costly cleaning of the processing machinery. An environmental concern of excessive finish application is increased Chemical Oxygen Demand (COD) on water supplies. Hence, the control and monitoring of the finish level on the synthetic fiber is critical to maximum productivity and processability in textile manufacturing processes.

There are several methods that can be used to measure the FOF of synthetic fibers, but the traditional methods for nylon fibers consist of the extraction of the finish oils from the fiber and the subsequent quantitative determination of the amount of finish present on a known weight of fiber (FOF). The finish oils are extracted with either hot or cold organic solvents (i.e., CCl_4) from the fiber. Two methods of determining the FOF are a gravimetric and an infrared (IR) method. In the gravimetric methods, the solvent is vaporized by either heat or steam, and the finish residue is weighed. The FOF is the weight of the finish residue compared to the initial sample weight. The gravimetric method is time consuming and may be influenced by the evaporation of volatile oils present in the finish during the vaporizing procedure. A more common FOF method consists of measuring the IR absorption of the various CH species in the extracted finish oils at 3.3 to 3.5 μm wavelength, and this absorption is compared to a known standard to obtain the FOF. Although both methods are accurate, they do require solvent extraction of the finish from the fiber and careful, extensive laboratory and calibration procedures. NIR methodology can be used to perform the analysis in less than 5 min without laborious extraction, laboratory, and calibration procedures.

A set of nylon66 spun yarn samples was analyzed using the NIR method and the IR solvent extraction method, using a filter-based NIR analyzer [17]. This example demonstrates a unique capability of the NIR technique — the ability to perform nondestructive multiple measurements on a single sample simultaneously. Here, both percent moisture and FOF could be measured on the same 3.5 g sample. A four-wavelength NIR Multiple Linear Regression (MLR) model consisting of 546, 1818, 2230, and 2270 nm was used in the FOF evaluation. As shown in Table 25.6, there was good agreement between the IR extraction and NIR methods, with a SEP of 0.04% FOF. The 546 nm wavelength can be attributed to the finish base, the 2230 and 2270 nm wavelengths are due to overtone and combination CH stretches from the various CH groups in the finish oils, and the 1818 nm wavelength is due to a $\text{C}=\text{O}/\text{O}-\text{H}$ stretch combination.

Similar results for FOF on nylon fibers were obtained with a scanning NIR analyzer, using a fiber optic remote sensing head [50]. The scanning unit offers the ability to utilize advanced chemometric modeling techniques — specifically, derivative math, Partial Least Squares (PLS), and Principal Component Regression (PCR) analyses. The samples were scanned over the 1100 to 2500 nm spectral range, and distinct second derivative spectral differences with varying FOF levels

TABLE 25.6
Finish-on-Fiber Comparison (%)

Item	IR	NIRA
X	1.06	1.12
Sx	0.032	0.048
SEP	—	0.04
95% CI	0.09	0.03

10

Source: Adapted from Dorrity and Tincher et al., *Textile Res. J.*, 65: 583 (1995).

were observed. A derivative PLS NIR calibration (1100 to 2300 nm spectral range) was developed, yielding excellent predictive results on a small verification sample set.

25.3.4 HEATSET TEMPERATURE MEASUREMENTS

Heatsetting, or twist setting, of synthetic yarns is one of the most critical processes in textile manufacturing. Heatsetting is a heat-treating process where heat is uniformly applied to the yarn under specific conditions of temperature, yarn tension, moisture, and heating time in order to obtain satisfactory fabric aesthetics, properties, and quality [51,52,53]. Heat treating the yarn is normally performed for stabilization against shrinkage, relaxation of internal stresses, bulk development, dye fixation, and twist setting or permanence. It gives the yarn a thermal and morphological “memory.” Heat has a significant impact on the molecular structure of synthetic yarns, which in turn influences the yarns’ chemical and physical properties and the morphological response of the yarns to further manufacturing steps. Type and quality of heatsetting can significantly affect the dye level, dyeing strike rate, dye bath exhaustion, dyeing variability, bulk appearance, and dye fastness of the yarn and resulting fabric.

Heatsetting consistency and uniformity are of critical importance in carpet yarn manufacturing. Nonuniform and inconsistent heatsetting can result in streaks in the finished carpet. A streak is a visual pattern parallel to the tufting direction that is caused by physical, optical, or dye difference, either within the yarn tuft-line or between adjacent yarn tuft-lines. One of the major components of heatsetting is the temperature at which the yarn is heatset — the heatset temperature. Several studies in carpet yarn manufacturing have shown that a small, but significant, variation in the heatset temperature (normally, $\pm 5^\circ$ from target) of the synthetic yarn during heatsetting can result in streaks in the finished carpet [54,55].

Two forms of heatsetting are primarily used in carpet yarn manufacturing — the hank process (Autoclave) and the continuous processes [47]. The autoclave process is labor-intensive, requires multiple processing steps, and is subject to heatsetting variations and nonuniformity. The continuous processes are single-step, require little labor, and are more attractive from both a technical and economic standpoint. Since the early-1970s, the major continuous heatsetting processes are the Suessen and Superba heatsetting systems. The Suessen is a dry method, while the Superba is a wet method. In the Superba process, the yarn is heatset with saturated steam; in the Suessen process, the yarn is heatset with dry heat, in which the small amount of steam present serves as a conditioning medium and to reduce oxidation and yarn discoloration. Superba and Suessen heatsetting result in yarns with vastly different morphological and dye properties [55–57]. Moisture lowers the glass transition temperature of the fiber and favors growth of large crystallites. Dry heat favors initial nucleation of the fiber structure, forming small crystallites up to a given temperature, where crystal

growth is favored. The result is that wet heatset yarns have more open structure and increased crystallinity than dry heatset yarns, which leads to an increased dyeing rate for the Superba yarns.

For hydrophilic yarns, like nylon66 and nylon6, the differences in the dyeability of the Superba and Suessen heatset yarns can be dramatic. Carpet yarns heatset by continuous processes are susceptible to streakiness due, in part, to (a) the large number of settings and controls that can deviate from standard conditions by operator choice or by normal system fluctuations and (b) the more uniform yarn produced by these systems can present a more perfect background for yarn defects to appear in the carpets. Thus, proper control and monitoring of the major parameters of the heatset process, especially the heatset temperature, is critical to the quality of the finished carpet.

Various techniques, such as x-ray diffraction, fiber density, dyeing or dye rate methods, and thermal methods of analysis, used to estimate the heat history and heatset temperature of synthetic yarns are not suitable for routine quality control/process control because of the time and extensive sample preparation required to perform the test. Therefore, NIR methodology offers a simple, rapid, and accurate technique for measuring the heatset temperature of carpet yarns.

Several authors have demonstrated the NIR method's superior capabilities in the measurement of nylon carpet yarn heatset temperature [14–18]. A typical NIR spectrum for Nylon is shown in Figure 25.13. Examples, using both filter-based and scanning NIR analyzers, will be presented for Superba and Suessen heatset, undyed (greige) yarns and for dyed yarns. The calibration samples ranged over a very wide (50°) range for both the Superba and Suessen models. Although several NIR prediction models have been proposed and utilized for the Superba and Suessen heatset Nylon carpet yarns, they contain two consistent features. First, the 2130 nm wavelength is common to many of the published models. Second, the wavelengths used in the prediction model normally occur in the same NIR spectral region from 1960 to 2310 nm. The prediction models are not necessarily “universal” models; rather, they are explicit models that will predict the heatset temperature differences within a given set or type of samples. Sample preparation and presentation, environmental, instrumental, and manufacturing process differences between different locations or mills may result in slight differences in the absolute prediction models for each location or mill. Many of these differences can be easily corrected with a slight bias change for each location or mill.

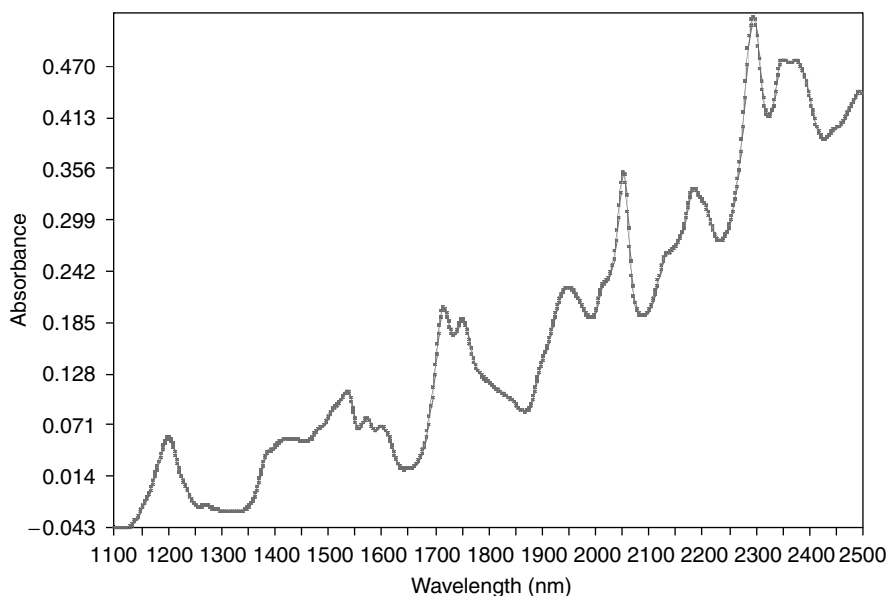


FIGURE 25.13 NIR spectra of heatset nylon yarn, 1100 to 2500 nm. [Redrawn from information in Rodgers and Ghosh, *Textile Res. J.*, 70: 519 (2000).]

Many published filter-based NIR prediction models contain the 2100, 2130, and 2270 nm wavelengths. Studies of the various NIR spectra obtained from various Suessen heatset nylon66 carpet yarns and comparison of the NIR spectra to other polymers with similar functional groups have yielded much information on probable wavelength assignments and the probable rationale for the 2130 nm wavelength commonality in the published NIR prediction equations [15]. The 2130 nm wavelength is due to both the C—H stretch/C=O stretch combination and N—H deformation/N—H stretch combination; the 2100 nm wavelength is due to C=O stretch combinations; and the 2270 nm wavelength is due to various CH and CH₂ stretches, bends, and deformations. It has been proposed that a consistent “slope” change is occurring at the 2100 and 2130 nm wavelengths, with increasing absorption at 2100 nm and decreasing absorption at 2130 nm with increasing heatset temperature. This slope change is most likely the driving force enabling the NIR method to measure the heatset temperature. The 2270 nm wavelength is most likely a reference or baseline band for the NIR spectra, as nonsignificant absorption differences were observed with increasing heatset temperature.

In addition to the above spectroscopic approach, the effects of changes in heat history on the NIR spectra, and on the prediction models for heatset temperature, of nylon carpet yarns have been studied from a crystallographic approach [16]. Various crystalline, thermal, and laser light scattering parameters were measured. It has been shown that the NIR spectrum and prediction models were responding to changes in the nylon fiber’s crystalline perfection with changing heatset temperature. The two components of the crystalline perfection index for nylon66 are the hydrogen bonding plane (d100) and the Van der Waals plane (d010). The d010 spacing, after 190°C, correlated well to increasing heatset temperature, while the d100 spacing exhibited a random pattern. Thus, the NIR spectrum and prediction models for heatset nylon66 carpet yarns were responding to changes in the C—H stretches in the Van der Waal’s plane.

25.3.4.1 Heatset Temperature Prediction

25.3.4.1.1 Greige Nylon Carpet Yarns

For Superba heatset carpet yarns, there was excellent agreement between the Superba actual measured process heatset temperature and the NIR predicted heatset temperatures for both nylon66 and 6 yarns [17]. Figure 25.14 is a comparison of a nylon66 Superba heatset yarn lot that was heatset over a wide temperature range, and it demonstrates the excellent agreement between the NIR method and the process heatset temperature. The multiple coefficients of determination (R^2) and SEP were very good, being .99 and 1.8°F, respectively.

A majority of the work to date has been on Suessen heatset nylon66 or 6 greige carpet yarns, where the NIR prediction model measured the Suessen heatset temperature of samples that had been Suessen heatset at different temperatures. Table 25.7 shows the NIR method’s capabilities for nine nylon66 sample sets. The NIR method did accurately measure the effective heatset temperature of these Suessen samples [17]. The NIR method and measurement variability were very good and comparable to the thermal methods of analysis, with an R^2 of .97 and pooled SEP of 2.3°C over the entire measurement range. These results were well within the temperature limits that are normally required to prevent dye streaks in the finished carpets. In many cases, it was observed that the NIR model’s and thermal analysis method’s prediction capability decreased below 190°C and above 220°C. This decrease resulted in higher-than-normal SEPs for the sample sets that contain the very low and very high heatset temperatures. However, the Suessen heatset temperature of primary interest to the carpet manufacturer normally lies in the region between 190 and 210°C. Within this region, the NIR method’s agreement with the Suessen heatset temperature was excellent, with an R^2 of .97 and pooled SEP of 1.2°C.

The deviation from linearity in the NIR model’s prediction capability for nylon66 yarns below 190°C was investigated with various crystalline, thermal, and laser light scattering parameters [16]. These deviations were normally small to moderate (<5°C) down to 185°C, but they consistently

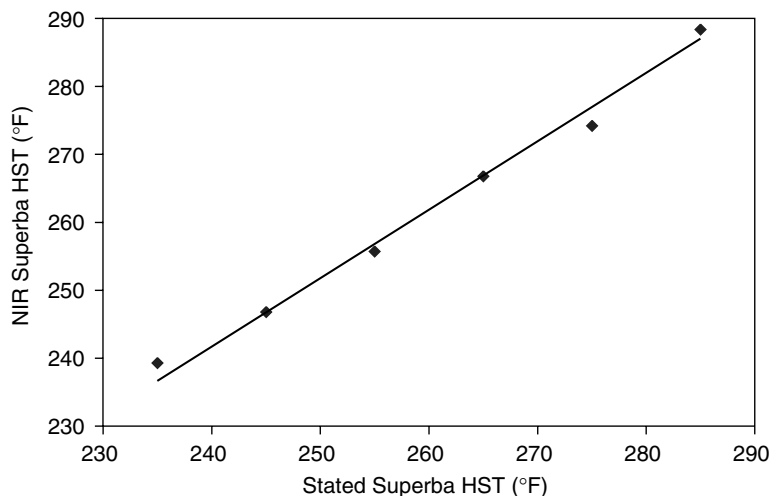


FIGURE 25.14 The relationship between Superba heatset temperature and the NIR method for nylon66 Greige yarn. (Redrawn from information in Rodgers, *Proc. 7th International NIRA Symposium*, Technicon Technical Center, Tarrytown, NY, 1984.)

TABLE 25.7
Suessen Heat-Set Temperature of Greige Nylon66 Yarns

HST range (°C) ^a	SEP (°C)			
	All		190–210°C	
	NIRA	Thermal	NIRA	Thermal
196–211	1.2	1.4	1.2	1.4
205–215	0.6	2.3	0.6	1.3
190–210	0.9	1.4	0.9	1.4
190–210	1.9	1.4	1.9	1.4
178–210	2.6	7.9	0.6	3.4
176–210	2.4	4.8	1.6	2.1
174–210	3.2	—	1.3	—
184–224	2.4	1.9	1.3	1.7
180–210	4.2	3.8	1.1	3.4
Pooled SEP (°C)	2.3	3.7	1.2	2.2

^a HST = heat set temperature.

appeared. Strong similarities were observed in the plots, vs. Suessen heatset temperature, of the NIR measured heatset temperature, thermal parameter phr (peak height ratio, the ratio of the peak heights from the two peaks obtained in a Differential Scanning Calorimeter [DSC] scan of heatset nylon66 yarns), laser light scattering intensity, and x-ray diffraction d010 spacing. For all of these methods, a definite change in the morphological response as a function of heatset temperature was exhibited in the plots by a marked change in slope at about 190°C. The distinct crystalline response transition occurring in nylon66 at about 190°C is the Brill transition [58–60]. Thus, the slight deviation in linearity in the NIR model's response below heatset temperatures of 190°C was primarily due to the Brill transition.

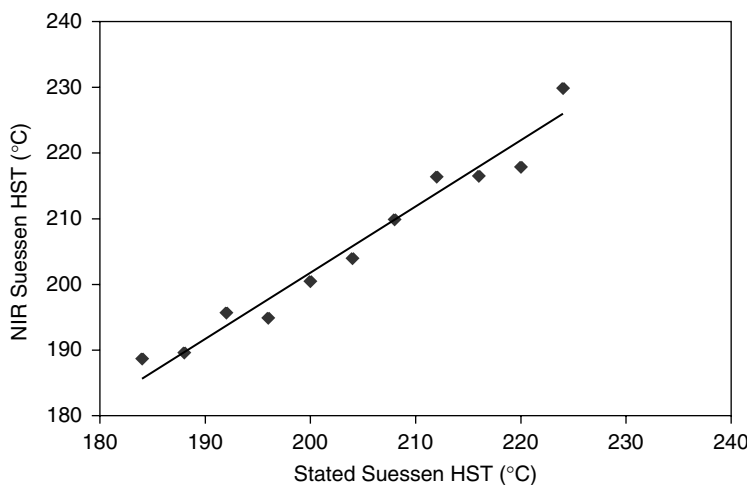


FIGURE 25.15 The relationship between Suessen heatset temperature and the NIR method for dyed nylon66 yarn. (Redrawn from information in Rodgers, *Proc. 7th International NIR Symposium*, Technicon Technical Center, Tarrytown, NY, 1984.)

25.3.4.1.2 Dyed Nylon Carpet Yarns

The difference in dyed vs. greige nylon carpet yarns is that the dyed yarns have undergone the additional textile manufacturing processes of tufting, dyeing, and finishing. These additional processes can impact the yarns' morphological properties, so it was by no means certain that the greige yarn calibration equations would perform satisfactorily for the dyed yarns. A comparative study is presented in which several samples of nylon66 yarns that had been Suessen heatset at different temperatures were dyed by several different laboratory and manufacturing processes. The NIR method, with only minor modifications to the three-wavelength MLR prediction equation, did accurately measure the effective heatset temperature of the dyed samples [17]. The method and measurement variability were good. Figure 25.15 shows this good agreement, on a nylon66 lot, with an R^2 of .96 and a SEP of 2.7°C. As observed with the greige yarn nylon66 samples, the model's prediction capability decreased below 190°C and above 220°C. The SEP within the 190 to 210°C range, the Suessen temperature region of primary interest to most nylon carpet yarn manufacturers, was 2.3°C.

25.3.4.2 Process Control Measurements, Nylon Carpet Yarns

In the previously mentioned heatset temperature examples, the work centered on "model-prediction" evaluations, where the NIR method measured the heatset temperature of nylon66 and 6 carpet yarns that had been heatset at different temperatures. The second broad area of heatset temperature measurements is "process control" evaluations, in which the NIR methodology monitors, on a routine basis, the heatset temperature differences on routine heatset samples. The process control measurement of heatset temperature will be illustrated using Suessen heatset nylon66 carpet yarns. Many carpet yarn manufacturers attempt to control the Suessen heatset temperature to within $\pm 3^\circ\text{C}$ of target (normally, 200°C) using the Suessen system controllers. As mentioned previously, heatset temperature values outside of $\pm 5^\circ\text{C}$ from target are of primary concern for streak propensity [55]. Therefore, NIR methodology can be used as a diagnostic tool to (a) monitor and identify off-temperature Suessen tunnels that have significantly low (or high) heatset temperatures (outlier/exception mode) and (b) monitor the effective Suessen heatset temperature over time of the Suessen unit (time monitoring mode). In both cases, the NIR methodology is used to *monitor* the Suessen unit for off-standard conditions.

TABLE 25.8
NIRA Measurement of Suessen Heat-Set Temperature, Routine Production^a

Day	NIRA, by shift (°C)			NIRA, daily (°C)
	S1	S2	S3	S
1	201	202	202	202
2	200	199	201	200
3	197	198	201	199
4	199	201	201	200
5	197	201	199	199
6	201	200	202	201
7	199	202	201	201
8	200	201	202	201
9	202	200	197	200
10	199	200	199	199
11	200	201	199	200
12	200	200	201	200
13	198	199	200	199
14	201	200	201	201
15	203	200	201	201
16	202	200	201	201
17	201	198	201	200
X	200	200	201	
Sx	1.7	1.1	1.3	
X		200		
Sxp ^b	1.4			

^a Target = 200°C for Suessen unit.

^b Sxp = pooled Sx, all shifts.

Source: Adapted from Rodgers and Ghosh, *Eastern Analytical Symposium*, Somerset, NJ, 1995.

A three-wavelength NIR prediction equation was used to monitor two large sets of routine process nylon66 Suessen heatset samples [61]. A total of 528 samples representing 88 work-shifts over a 30-day period were evaluated. The NIR method successfully measured the Suessen heatset temperature of both lots of routine production, as shown for the second lot in Table 25.8. No NIR values, shift or daily average, were outside the $\pm 5^\circ\text{C}$ “streak limit,” indicating excellent process control monitoring. The NIR method predicted that no Suessen heatsetting streaks should occur in the finished carpets. The NIR method also detected the known slight heatset temperature differences between the six tunnels in a Suessen unit. The yarns from both lots were tufted, dyed, and finished into carpets. As the NIR method had predicted, no streaks were observed in the carpets of both lots.

25.3.5 TEXTILE PRODUCT IDENTIFICATION

Identification of different polymers, their properties, and their morphological differences normally requires extensive testing and complicated time-consuming analysis. Many statistical analysis methods have been combined with NIR to identify dissimilar textile products, including PCA, linear or multiple regression, PLS, derivative math, Mahalanobis distance (i.e., Discriminate Analysis), nearest neighbors techniques, and neural networks. Most textile fibers, yarns, and fabrics have chemical structures that yield complex NIR spectra, and as such these species normally require three or

more wavelengths to classify the material. The following examples are given to demonstrate the potential and high flexibility of the various identification software programs.

25.3.5.1 Nylon and Heatsetting Type Identification

It is well known that different nylon polymer types (nylon66 and 6) and different heatsetting methods (Suessen and Superba) will have different fiber morphology and molecular structure, and these differences can often result in dye streaks if yarns of different heatsetting or nylon polymer type are mixed. The NIR method, using both filter-based and scanning NIR analyzers with the Discriminate Analysis technique, has been shown to quickly and accurately identify nylon66 and 6 yarns that have been either Suessen or Superba heatset [26]. The Discriminate Analysis technique and theory have been previously discussed [62]. The Discriminate Analysis technique develops calibration identification equations that can be used to identify samples that are similar to the calibration samples. The evaluation consisted of three separate identification sets — nylon polymer type, heatsetting type, and specific identity. For the nylon polymer-type and heatsetting-type identification, the Discriminate Analysis software was being asked to identify the standard one variable (polymer type or heatsetting type) based on the detection of the known morphological differences between nylon66 and 6 yarns and between Suessen and Superba heatset yarns. The specific-identity classification presented a much more complex problem to the qualitative software, for it was being asked to identify a sample based on two variables — polymer type *and* heatsetting type. Three- and four-wavelength calibration equations were obtained for the identification sets. The Mahalanobis distances between the identification sets were large. The larger the Mahalanobis distances between the groups, the higher the probability for the Discriminate Analysis technique to correctly classify the sample sets. The size of the Mahalanobis distances between groups indicated that the molecular structure differences measured by the Discriminate Analysis technique were greater for the polymer types than for the heatsetting types. The calibration equations were then evaluated for predictability using four sample sets of 34 yarns. The 34 unknowns were classified according to polymer type, heatsetting type, and specific identity (Table 25.9). The classification of the unknowns was excellent, with 100% correct identification for all three calibrations.

25.3.5.2 Polyester Fiber by Different Producers

It has been observed that polyester staple fibers that are of different tenacity (response to tensile stress placed on the fiber) levels and that are produced by different fiber producers will have different fabric dyeing and manufacturing production criteria. Both regular and high-tenacity staple polyester fiber were obtained from four fiber producers [27]. The NIR method, using both filter-based and scanning analyzers with Discriminate Analysis software, successfully identified and classified the staple polyester samples by producer, by tenacity level, and simultaneously by producer and tenacity (Table 25.10). The fiber identification was successful even when the polyester fiber was blended with cotton in a 50/50 blend. The eight groups of fibers were dyed. Variations in dye uptake, which would have resulted in quality problems, were present between the various tenacity levels and producers. This NIR method provided a quick and accurate technique for the identification of the polyester fiber before quality problems could occur.

25.3.5.3 Fiber/Polymer-Type Identification

Qualitative analyses can be used to identify and classify fibers of dissimilar or different polymeric type. Often, the spectral differences between the different fiber types can be quite small and minor (i.e., nylon66 vs. nylon6). Several qualitative measurements for fiber identification have been reported, but most of these use the more traditional spectral analysis and spectral pattern recognition

TABLE 25.9
Discriminant Analysis Calibration Equations

Calibration set	Wavelengths (nm)	Mahalanobis distances (MD)		
		From	To	MD
Nylon polymer type	2472	Nylon66	Nylon6	65.60
	2486			
	2500			
Heat-setting type	1884	Suessen	Superba	6.96
	2094			
	2122			
	2500			
Specific identity	1506	Suessen/nylon66	Suessen/nylon6	45.44
	1534		Superba/nylon66	11.11
	2472	Suessen/nylon6	Superba/nylon6	37.15
	2500		Superba/nylon66	56.28
			Superba/nylon6	8.38
			Superba/nylon66	47.97

Source: Adapted from Rodgers and Ghosh, *Textile Res. J.*, 71: 135 (2001).

TABLE 25.10
Identification of Nylon Heat-Set Unknowns

Sample set	Number	Unknowns		Correct classification		
		Polymer type	Heat-setting type	Polymer type	Heat-setting type	Specific identity
A	10	Nylon66	Suessen	10	10	10
B	10	Nylon6	Suessen	10	10	10
C	4	Nylon66	Superba	4	4	4
D	10	Nylon6	Superba	10	10	10

Source: Adapted from Rodgers and Ghosh, *Textile Res. J.*, 71: 135 (2001).

techniques [25–28,30,31]. A novel approach to fiber identification was the application of neural networks to the spectral analysis and spectral pattern recognition programs [29]. A scanning NIR unit measured the spectra of 390 synthetic and natural fiber samples from 17 different fiber types/groups. Using backpropagation for the NIR network learning rule and mean subtraction, a neural network was developed that included 700 input absorbance wavelengths, 21 hidden nodes, and 17 output nodes. The neural network not only successfully identified fibers from the 17 fiber groups, it correctly identified the fiber components in blends (i.e., linen-rayon blend) from fabrics that had not been included in the training sets.

25.3.5.4 Carpet Fiber Identification for Recycling

Recycling programs are growing in popularity and importance, especially for the carpet industry. A key component of a potential carpet recycling program is the rapid and accurate identification of the

fiber/polymer type used in the carpet face fiber in the returned carpet. Several fiber producers have developed their own in-house NIR carpet fiber identification programs, and several NIR analyzers specifically designed to identify the carpet face fiber are commercially available. An example of the flexibility of the NIR technique for carpet face fiber identification can be shown by evaluations performed with one of the first commercially available units — the Carpet Analyzer [28]. The Carpet Analyzer is a mobile, fiber-optic probe, scanning NIR unit that scans the sample quickly (5 s or less analysis time) from 1200 to 2400 nm. Derivative PCA calibrations and an internal fiber identification library were developed to accurately identify the polymer type of most carpet face fibers, including nylon66, nylon6, polyester, polypropylene, wool, and acrylic. The hand-held probe is placed against the carpet surface, the NIR spectrum collected, and the carpet spectrum is compared to the fiber types present in the pre-calibrated internal library. Using similar procedures established and demonstrated by the Carpet Analyzer, many other commercial NIR/FT-NIR analyzers are available today for the identification of carpet face fiber.

Investigations were also conducted to obtain a better understanding of the impact of key carpet component and instrumental variables on the identification of carpet face fiber by NIR [31]. The improved understanding of these influences on the NIR calibrations could explain why some NIR identification models appear to be more robust and “capable” than others. Static (load sample into a sample cell or place the same against the NIR unit’s port) and fiber-optic probe analyses were performed on an extensive series of carpets and yarns of nylon66, nylon6, polypropylene, polyester, and acrylic polymer. The carpet components evaluated were fiber type, color, carpet and yarn construction, and dyeing/coloration method. The instrumental variables evaluated were sampling system (static vs. probe) and measurement speed (scans per sample). As expected, distinct spectral differences were observed between the nylons, polyester, polypropylene, and acrylic fibers (Figure 25.16). These differences were accentuated with the use of second derivative spectra, and the second derivative spectra were able to differentiate between nylon66 and nylon6 (Figure 25.17). Color and construction differences did result in changes in the intensity of the NIR absorbance and, in some instances, noticeably and distinct shifts in the spectral baseline (i.e., pigments and type carpet surface). However, while spectral changes were observed with moderate

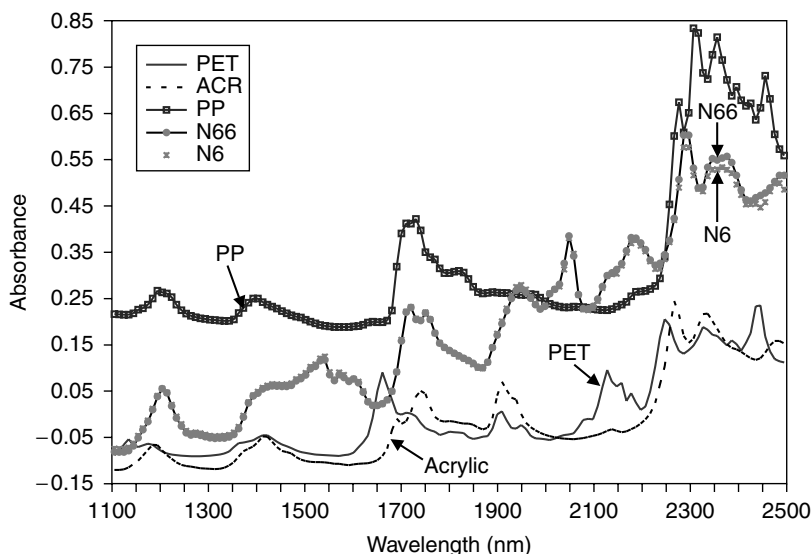


FIGURE 25.16 NIR absorbance spectra for polyester (PET), acrylic (ACR), polypropylene (PP), nylon66 (N66), and nylon6 (N6) fiber types, 1000 to 2500 nm. [Redrawn from information in Rodgers, *AATCC Review*, 2: 27 (2002).]

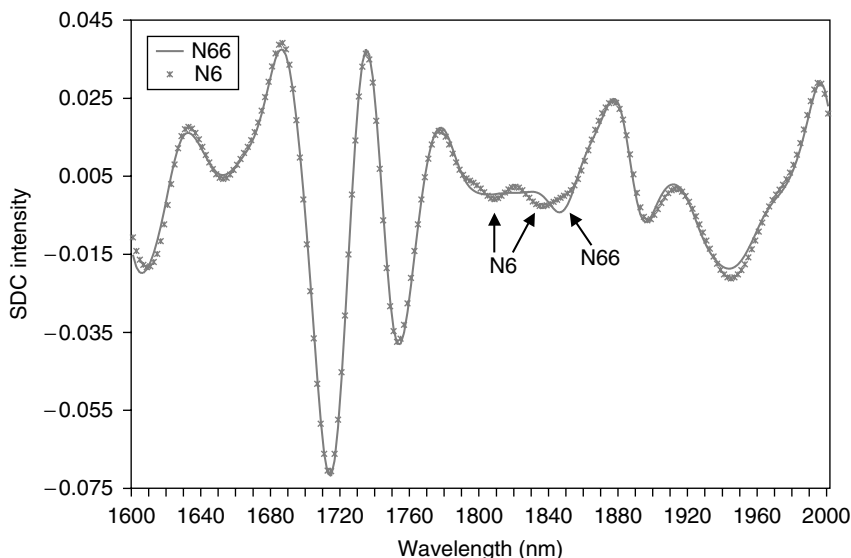


FIGURE 25.17 NIR second derivative spectra for nylon66 (N66) and nylon6 (N6) fiber types, 1600 to 2000 nm. [Redrawn from information in Rodgers, *AATCC Review*, 2: 27 (2002).]

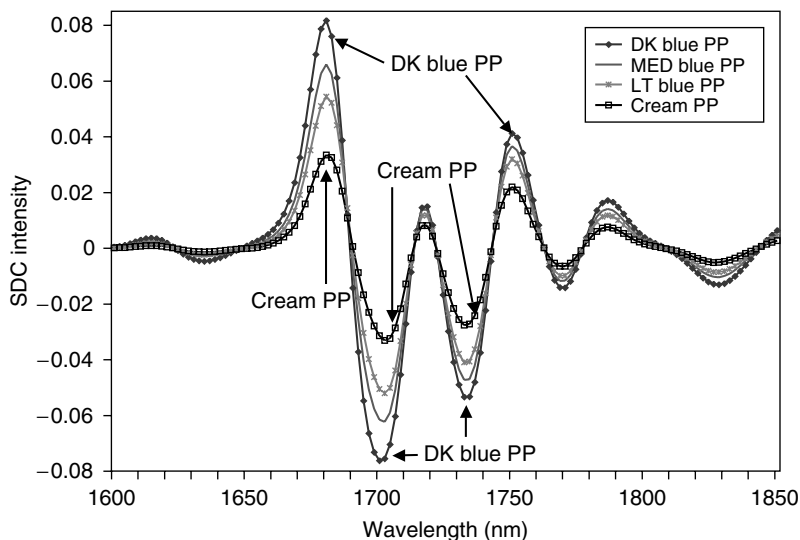


FIGURE 25.18 NIR second derivative spectra for different pigment color polypropylene (PP) fiber, 1600 to 1850 nm. [Redrawn from information in Rodgers, *AATCC Review*, 2: 27 (2002).]

to large changes in the carpet components, second derivative spectra demonstrated that shifts in the polymer type's characteristic wavelengths did not occur (Figure 25.18 for an example). The major instrumental parameter to influence the spectral curve was the sampling system — static vs. probe. Only minor spectral baseline movements were observed with changes in analysis time (number of scans per sample). Significant spectral differences were observed in the spectral intensities of the polymers between the static and probe systems, with the probe yielding lower absorbance differences (peak-to-baseline amplitude) (Figure 25.19). However, once again, the second derivative

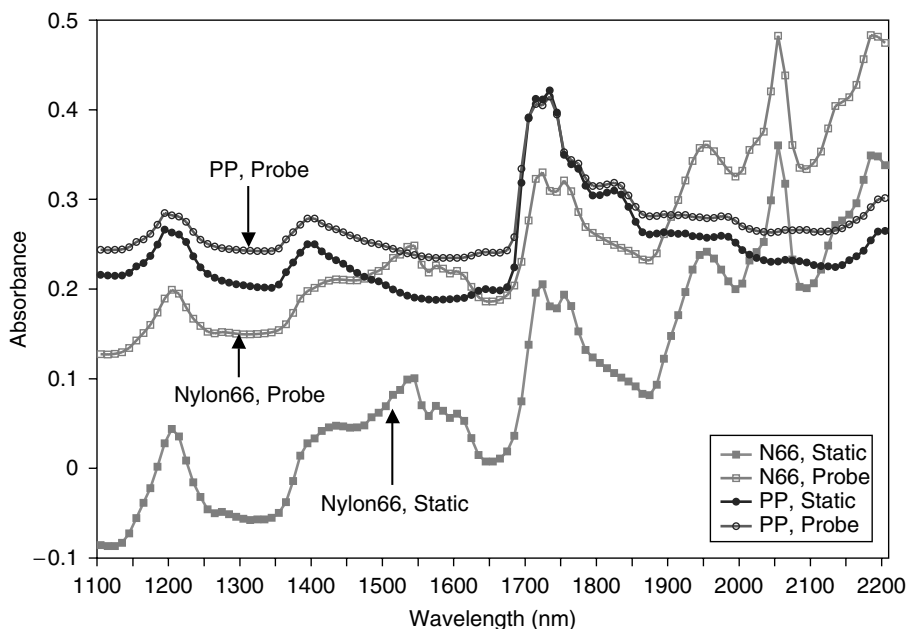


FIGURE 25.19 NIR absorbance spectra for polypropylene (PP) and nylon66 (N66) fiber types, static/bench-top vs. fiber-optic probe sampling systems, 1000 to 2200 nm. [Redrawn from information in Rodgers, *AATCC Review*, 2: 27 (2002).]

TABLE 25.11
Identification of Staple Polyester Tenacity Unknowns

Sample set	Unknowns			Correct classification		
	Number	Fiber producer	Tenacity level	Fiber producer	Tenacity level	Simultaneous
1	5	DuPont	High	5	5	5
2	5	DuPont	Regular	5	5	5
3	5	Celanese	High	5	5	5
4	5	Celanese	Regular	5	5	5
5	5	Eastman	High	5	5	5
6	5	Eastman	Regular	5	5	5
7	5	Hoeschst	High	5	5	5
8	5	Hoeschst	Regular	5	5	5

Source: Adapted from Ghosh, *J. Textile Institute*, 84: 85 (1993).

spectra demonstrated that shifts in the polymer type's characteristic wavelengths did not occur as a result of these instrumental parameters. Thus, for a robustly calibrated system (many carpet components), carpet component and instrumental parameters resulted in minimal effects on fiber-type identification. Optimal polymer-type identification was obtained with second derivative NIR calibrations. The NIR calibrations were validated on almost 200 carpet and fiber samples, with 100% correct identification rates using the second derivative calibration (Table 25.11).

25.3.6 AT-LINE/ON-LINE ANALYSES

The advent of stringent financial and cost controls, manufacturing efficiencies and productivity, and “re-engineering” have placed strong (and increasing) pressure on laboratory analyses and laboratory resources. Although these financial and productivity pressures tend to restrict the number and usage of laboratory analyses, the need for “just-in-time,” accurate, and precise analytical results on key physical and chemical properties of a product is of an even greater importance today. One answer to this need for fast, precise, and accurate analytical results in today’s business environment is the use of at-line and on-line analyses. At-line analyses are those in which the NIR analyzer is located in the manufacturing area, and the samples are brought to the analyzer for analysis by area operators rather than sending the sample to the lab. On-line analyses are those in which the NIR analyzer is placed in the manufacturing process, and the NIR measurement, sample handling, and sample preparation are controlled automatically by a computer. Recent instrumental advances by manufacturers of NIR instrumentation have resulted in an increasing use of NIR reflectance measurements on textile substrates, both at-line and on-line.

There are several procedures that can be utilized to evaluate and implement at-line and on-line measurements. One of the more common NIR analysis processes consists of three phases — bench-top unit evaluation, pilot plant or laboratory evaluation of the at-line/on-line instrumentation, and manufacturing area evaluation. The Bench-Top phase establishes the method definition and analysis potential. The Pilot Plant phase verifies the method and instrumentation for the analysis under consideration, determines the optimum sampling system, and identifies operational parameters and concerns. The Manufacturing area phase optimizes data treatment, optimizes operational and instrumental parameters, and proceeds with method implementation.

At-line and on-line NIR measurements present unique complexities and operational concerns [34–37]. Several major operational concerns for at-line and on-line measurements of textiles must be considered and planned for before NIR evaluations. The major operational concerns are as follows:

- Sample matrix effects
- Type of data treatment
- Temperature and environmental impacts
- Instrumental parameters
- Details of the experimental process.

Sample matrix effects can include the number of scans required to obtain the desired accuracy, precision, and analysis time; the number of readings or packings of the sample; sample presentation and conformity (bobbin wind, fabric weave, surface hardness, etc.); fabric surface effects (additives, denier, fiber cross section, etc.); and where on the textile surface to measure the sample. Data treatment parameters include the type of mathematical or chemometric model (MLR, PLS, PCR, derivative, etc.) and type and strategic emphasis of statistical parameters (within-sample variability, between-sample variability, R^2 , residual analyses, etc.). Temperature and environmental parameters consist of the effect of process/measurement temperature on the NIR spectra, type of material (fiber type, hazards, etc.), chemical groups/environment present, and potential absorption deviations due to chemical or physical parameters present in the sample (water, additives, etc.). Instrumental parameters involve the spectral range, the size and weight of the sampling system, the expected use of the measurement (ergonomic effects, at-line or on-line, etc.), simplicity and ease of operation, and data manipulation and collation (averaging, automatic saving of spectra, background subtraction, etc.). The experimental process can include the quantity of process detail required to obtain an acceptable at-line/on-line NIR method, the quality and composition of the calibration and prediction sample sets (natural or prepared samples, accuracy and precision of reference method, sample selection techniques, etc.), the number of calibration and prediction

samples, and a defined end-state (conditions and requirements to be completed for a successful implementation).

Three examples — one at-line, one on-line, and one at-line/“pseudo” on-line — are presented to demonstrate the flexibility and versatility of the NIR spectroscopy and instrumentation for process measurements.

25.3.6.1 At-Line Measurement of Carpet Yarn

Heatset Temperature

The importance of the uniform and consistent application of heat/heat history during the heatsetting of carpet yarns has been described previously. The heatset temperature is one of the most important operating variables in the heatsetting process. Previous investigations described earlier emphasized Bench-Top, and Laboratory NIR measurements of heatset temperature. This project’s objective was to demonstrate the transference of the measurement of Suessen and Superba heatset temperature for nylon66 carpet yarns from Bench-Top to at-line and to provide manufacturing with a simple and rapid at-line NIR method to monitor the heatset temperature of Suessen heatset nylon carpet yarn *directly on the bobbin* [34–37]. The defined end-state was the development of a method with good at-line accuracy and precision (residual error/SEP of $<4^{\circ}\text{C}$ for Suessen yarns and $<4^{\circ}\text{F}$ for Superba yarns). The initial measurements were performed in the laboratory with a scanning NIR analyzer and two sampling systems — a fiber-optic probe (“Probe”) and a fiber-optic remote sensing head (“Head”) [36]. Second-derivative NIR spectra of the heatset yarns were compared, as a function of heatset temperature, for major areas of spectral difference (Figure 25.20). Spectral differences were observed primarily in the 1960 to 2200 nm region, and this spectral region was selected for calibration development. Excellent agreement was obtained between the stated heatset temperatures and the NIR-determined heatset temperatures for both sampling systems, with high R^2 s and low SEPs (Table 25.12). The Probe system was the most promising overall system for *at-line* heatset temperature measurements on the bobbin. The Probe

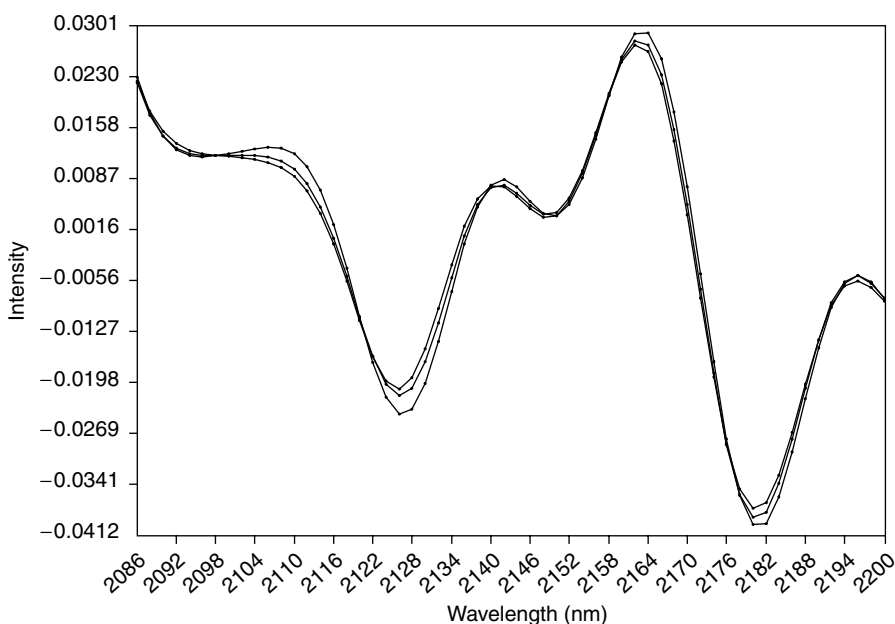


FIGURE 25.20 At-line second derivative NIR spectra of Suessen heatset nylon66 yarn, 2086 to 2200 nm. [Redrawn from information in Rodgers and Ghosh, *Textile Res. J.*, 70: 519 (2000).]

TABLE 25.12
At-Line NIR Measurements of HeatSet Temperature

Sampling system	Parameter	Validation set results	
Superba heatset		260–285°F	All
Probe ^a	Bias (°F)	−0.6	0.3
	R^2	.98	.99
	SEP (°F)	1.08	1.87
Head ^a	Bias (°F)	0.5	1.7
	R^2	.99	.99
	SEP (°F)	0.98	1.80
Suessen heatset		190–210°C	All
Probe ^b	Bias (°C)	−0.6	0.2
	R^2	.98	.99
	SEP (°C)	0.72	2.20
Head ^a	Bias (°C)	0.1	2.0
	R^2	.98	.98
	SEP (°C)	0.78	2.30

^a Second derivative MLR calibration equation.

^b Second derivative PLS calibration equation.

Source: Rodgers and Ghosh, *Textile Res. J.*, 70: 519 (2000).

TABLE 25.13
At-Line Determination of Production Lot Quality by NIR

Sample set	Production lot quality	SEP (°F)
Superba 1	Good	0.78
Superba 2	Problem	2.31
Superba 3	Unknown	2.03

Source: Rodgers and Ghosh, *Textile Res. J.*, 71: 135 (2001).

system yielded a simpler operation, lower analyzer weight, faster analysis time, increased flexibility, superior measurements on small/low weight bobbins, and acceptable accuracy. The Head system was best suited for *on-line* heatset temperature measurements (larger sampling area and greater precision).

The NIR analyzer, with Probe sampling system, was evaluated in the manufacturing environment at several carpet manufacturers [37]. Bobbin/at-line NIR calibrations were developed in the process area laboratory and transferred to manufacturing locations. Very good heatset temperature agreement was obtained for both wide-range and routine production samples, with low SEPs and high R^2 s. Quality differences in routine production bobbins from different lots of Superba heatset yarns were observed. On the basis of the SEP of each lot (consistency of heatset temperature for bobbins in a lot), the at-line NIR technique correctly identified and segregated the “good quality” lot from a “problem quality” lot at one manufacturer (Table 25.13). In addition, the NIR method determined that an “unknown quality” lot would be marginally acceptable for shipment — an assessment subsequently confirmed by the manufacturer. At another manufacturer, yarn product/heatset type

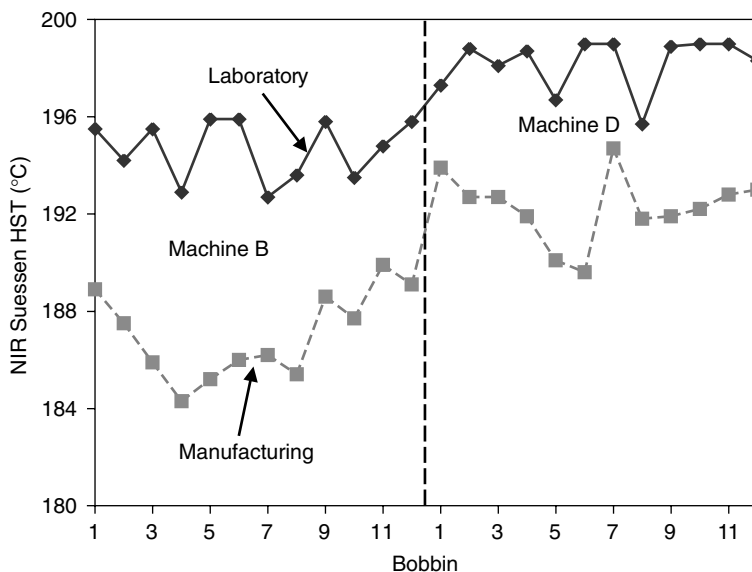


FIGURE 25.21 Effect of environmental measurement conditions on the NIR-determined Suessen heatset temperature, regular (Machine D) vs. hi-twist (Machine B) heatsetting machines. [Redrawn from information in Rodgers and Ghosh, *Textile Res. J.*, 71: 135 (2001).]

differences (regular Suessen vs. Hi-twist/stufferbox Suessen) and environmental influences were observed. Hi-twist heatset yarns yielded different diffuse reflectance than regular Suessen heatset yarns. The expected slight difference in NIR-determined heatset temperature was observed, but the precision for each yarn type was very similar — implying comparable product quality (Figure 25.21). The major environmental influence was area temperature — a 15 to 20°F temperature difference existed between the laboratory and manufacturing location measurements. As expected, significant shifts in NIR-determined heatset temperature were observed, but the precision for each yarn (SEP) remained excellent and comparable. Thus, only a slight bias adjustment to target would be required to use the laboratory at-line NIR calibration in the manufacturing area. The at-line NIR method provided a rapid, accurate, and precise process control/quality control measurement of nylon66 carpet yarn heatset temperature.

25.3.6.2 On-Line Measurement of PVA Size on Warp Yarns

Size is added to warp yarns to minimize warp end breaks during weaving. The amount of PVA and uniform application of size to the warp (size add-on) can significantly influence weaving efficiency. A common sizing agent used for PET blend warp yarns is polyvinyl alcohol (PVA). An on-line NIR method was developed that accurately and precisely monitored the PVA add-on on the warp sizing (slasher) machine [38]. NIR spectra were collected from 1100 to 2500 nm using a scanning NIR analyzer with a remote fiber-optic sensing head. Distinct second-derivative spectral differences were observed between 1700 to 1800 nm with increasing size add-on (Figure 25.22). Both derivative MLR and PLS NIR calibrations were developed and evaluated. The best NIR-calculated PVA add-on agreement was obtained with a derivative PLS model, with a R^2 of .99 and SEP of 0.23% size for laboratory slasher samples (Figure 25.23). The MLR models also performed well on the laboratory slasher, but a long-term drift was observed for samples prepared and measured on a commercial slasher (manufacturing environment). This long-term drift resulted in much greater SEPs for the MLR equations than observed for the PLS equation. Thus, the derivative PLS equation was the

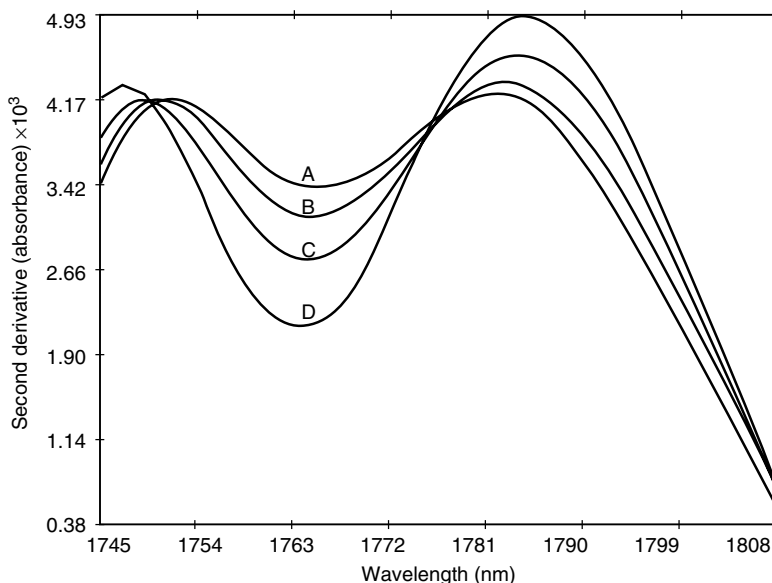


FIGURE 25.22 On-line second derivative NIR spectra of PVA-sized PET yarn, 1700 to 1800 nm. A = 9.0, B = 11.7, C = 14.6, D = 20.1 %Size. [Redrawn from information in Ghosh, *J. Textile Institute*, 84: 85 (1993).]

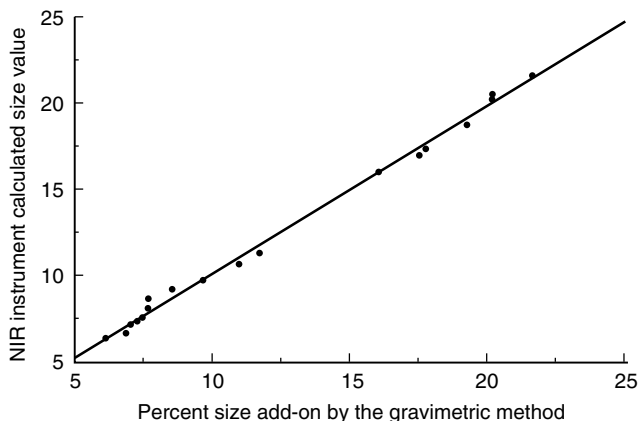


FIGURE 25.23 Relationship between laboratory analyzed percent size on warp and on-line NIR-calculated percent size using a PLS calibration. [Redrawn from information in Ghosh, *J. Textile Institute*, 84: 85 (1993).]

best and most stable NIR calibration for PVA add-on for commercial slashers. The PLS model was also evaluated in a process control mode. The PLS calibration correctly monitored the slasher as it was operating normally (only one outlier) and as it was operating during periods of process upsets and machine changes. The changes in warping efficiency and size add-on during these process upsets were observed.

25.3.6.3 At-Line/"Pseudo" On-Line Measurement of Tire and Carpet Yarn Moisture and Finish-on-Fiber

The importance of accurately and precisely measuring the moisture content and FOF applied to textile products has been described previously. The moisture content and FOF applied to the surface

of textile fibers can significantly influence the fiber's physical properties, manufacturing processes and control, quality, and productivity. Many investigations have emphasized specific moisture or FOF analyses using sophisticated bench-top or on-line NIR analyzers. One investigation studied the feasibility of using a portable, low-cost NIR analyzer to measure the moisture of textile bobbins in the manufacturing area, during preparation of the bobbins in spinning [33]. The non-contact measurement was performed in the spinning manufacturing area, and the NIR unit was moved from position to position down the manufacturing unit. Thus, this form of process analysis can be considered as "pseudo" on-line — an at-line analyzer measuring bobbins on each position sequentially.

The moisture content and the FOF applied to the fiber are often critical control variables, and they can significantly influence and affect the fiber's physical properties, manufacturing processes and quality, and productivity. A recurring difficulty and concern in the production of nylon yarn bobbins is the rapid detection of outlier or out-of-specification moisture and FOF bobbins and machine positions at or near the application of the finish in manufacturing. As the identification of moderate process deviations that would lead to outlier bobbins was the primary concern of the manufacturing area, the high degree of accuracy normally associated with a full scanning NIR analyzer was not required for this application. A portable, low-cost, noncontact, semi-quantitative NIR analyzer was used to measure moisture, directly on tire and carpet yarn bobbins, in various manufacturing areas (at-line). Moisture content calibrations were developed with bobbins of wide moisture ranges for tire yarn and carpet yarn products. The NIR unit successfully monitored moisture differences between bobbins in several manufacturing areas (spinning and lag areas, spun and drawn yarns, etc.) and in the laboratory after the bobbins had acclimated. Using the same NIR calibration for the manufacturing area and laboratory moisture analyses, very good method agreement was observed between the NIR and thermal/weight loss (reference) moisture results for the manufacturing area and laboratory analyses, well within the desired moisture agreement between the two techniques of $\pm 0.5\%$ moisture for over 90% of the samples analyzed (Table 25.14 for an example). Outlier bobbins and off-standard manufacturing positions were readily and quickly identified.

In spinning, it may be feasible to use the spun fiber moisture measurement as a surrogate measurement for bobbin FOF. For spun fiber production, FOF normally increases/decreases in the same direction as the moisture. A strong correlation was observed between on-the-bobbin

TABLE 25.14
At-Line Moisture, Portable NIR Unit, Laboratory, and Manufacturing Areas

Samples	Manufacturing area moisture (%)			Laboratory moisture (%) ^a		
	NIR	Brabender	Delta	NIR	Brabender	Delta
1	2.74	3.12	0.38	4.76	4.79	0.03
2	2.62	2.90	0.28	4.86	4.97	0.11
3	2.62	2.86	0.24	5.00	5.03	0.03
4	2.59	2.88	0.29	4.91	4.87	-0.04
5	2.14	2.27	0.13	4.81	4.75	-0.06
6	2.61	2.80	0.19	4.79	4.82	0.03
7	2.52	2.80	0.28	5.16	4.97	-0.19
8	2.62	2.68	0.06	5.01	5.02	0.01
AVG	2.56	2.79	0.23	4.91	4.90	-0.01
SD	0.31	0.31	0.30	0.21	0.17	0.13

^a Laboratory moisture measured after a 24 h lag.

Source: Rodgers, *AATCC Review*, 2: 27 (2002).

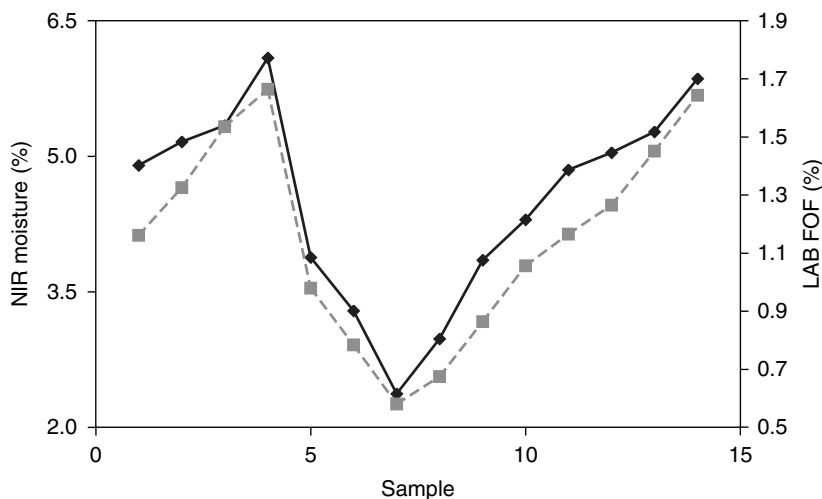


FIGURE 25.24 The relationship between laboratory and at-line NIR finish-on-fiber for nylon66 bobbins. (Redrawn from information in Rodgers, et al., *Pittsburgh Conference 2002*, New Orleans, LA, 2002.)

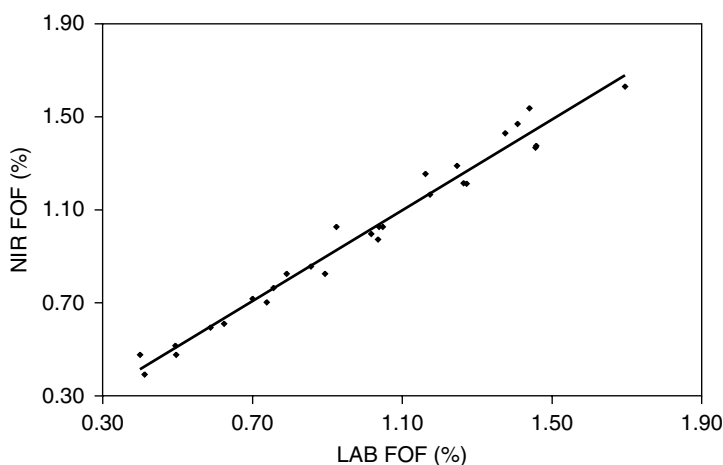


FIGURE 25.25 The relationship between laboratory and at-line NIR finish-on-fiber for nylon66 BCF carpet bobbins (wide range of finish-on-fiber). (Redrawn from information in Rodgers, et al., *Pittsburgh Conference 2002*, New Orleans, LA, 2002.)

moisture obtained by the NIR unit during spinning and FOF (Figure 25.24). Surrogate, direct-read FOF calibrations were developed for tire and carpet yarn products. The robustness of the NIR surrogate FOF calibrations was evaluated on bobbins from different products (different denier/size, tints/“color,” end uses, manufacturing conditions, plant locations, etc.). Very good method agreement was observed between the NIR and laboratory FOF results (Figure 25.25 and Figure 25.26). Of critical importance to the manufacturing area was the rapid at-line identification of several “outlier” bobbins of low FOF due to manufacturing upsets, preventing their contamination of downstream processes. The influences and effects of all yarn differences, manufacturing locations, and operational and environmental conditions were slight and manageable.

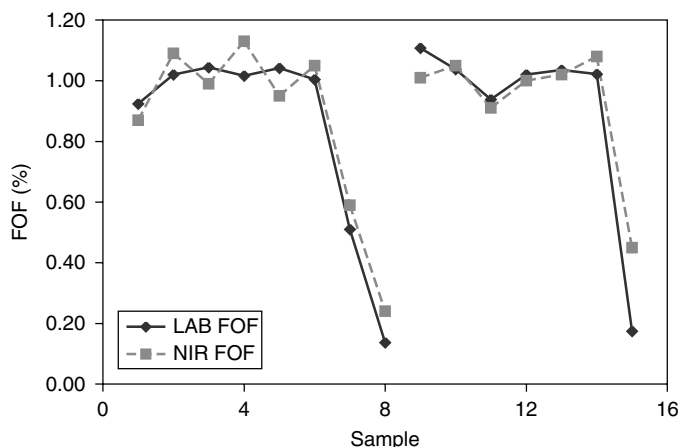


FIGURE 25.26 The relationship between laboratory and at-line NIR finish-on-fiber for nylon66 BCF carpet routine production bobbins. (Re-drawn from information in Rodgers, et al., *Pittsburgh Conference 2002*, New Orleans, LA, 2002.)

REFERENCES

1. H. H. Perkins, *Textile Ind-Atlanta*, 49: 141–145 (1977).
2. S. Ghosh and R. B. Roy, *JTI*, 79: 3 (1988).
3. R. Giangiacoma, et al., *Encyclopedia of Analytical Chemistry*, Ed. Robert A Meyers, John Wiley & Sons Ltd., Chester (1981).
4. R. Giangiacoma, et al., *J. Food Sci.*, 46: 531 (1981).
5. M. Meurens, *Proceedings of 3rd Annual Users Conference for NIR Research*, Pacific Scientific Inc., Silver Springs, Maryland, 1984.
6. E. Lanza and B. W. Li, *J. Food Sci.*, 49: 995 (1984).
7. D. E. Honigs, G. M. Hieftje and T. Hirschfeld, *Appl. Spectrosc.*, 38: 317 (1984).
8. S. M. Edelstein, *Am. Dyes Stuff Ref.*, 25: 458 (1936).
9. E. S. Olson, *Wet Processing, Volume I, Preparation of Fibers and Fabrics*, Noyes, Park Ridge, New Jersey, 1983, p. 164.
10. V. I. Nikitin, *Vestn. Leingr.*, 3: 33 (1950).
11. J. Gjønnes and N. Norman, *ACTA Chem. Scand.*, 14: 683 (1960).
12. J. D. Guthrie, C. L. Hoftpanir, M. F. Stansbury and W. A. Reeves, *Survey of the Chemical Composition of Cotton Fibers, Cotton Seed, Peanuts, and Sweet Potatoes, A Literature Review*, United States Department of Agriculture, Agricultural Research Administration, Bureau of Agriculture and Industrial Chemistry, AIC-61 (revised March 1949).
13. K. Ward, Jr., *Chemistry and Chemical Technology of Cotton*, Interscience, New York, 1955, p. 15.
14. S. Ghosh and J. Rodgers, *Textile Res. J.*, 55: 556 (1985).
15. J. Rodgers and S. Lee, *Textile Res. J.*, 59: 513 (1989).
16. J. Rodgers and S. Lee, *Textile Res. J.*, 61: 531 (1991).
17. J. Rodgers, Specific Applications of NIRA in the Textile Industry, *Proc. 7th International NIRA Symposium*, Technicon Technical Center, Tarrytown, NY, 1984.
18. W. Tincher, et al., *Textile Chem. Color*, 18: 25 (1986).
19. S. Ghosh, Applications of Heat History Measurements of Polyester Yarn, *Proc. 10th International NIRA Symposium*, Technicon Technical Center, Tarrytown, NY, 1986.
20. J. Dorrity, W. Tincher, et al., *Textile Res. J.*, 65: 583 (1995).
21. M. Drews and S. Harris, Using NIR to Check Latex Binders in Nonwovens, *AATCC NIR Technology in Textiles Symposium*, Charlotte, NC, 1990.
22. L. Lemere, *Textile Chem. Color*, 21: 19 (1989).
23. B. Hill, Evaluation of a NIR Measuring Process for Sewing Thread, *AATCC Textile Applications of NIR Technology Symposium*, Asheville, NC, 1993.

24. J. Harper, Applications of Near Infrared Reflectance Spectroscopy to Rayon Yarns, *AATCC Textile Applications of NIR Technology Symposium*, Asheville, NC, 1993.
25. H. Howell and J. Davis, *Textile Chem. Color*, 23: 69 (1991).
26. S. Ghosh and J. Rodgers, *Meiland Textilberichte*, 69: 361 (1988).
27. S. Ghosh, Applicable Textile Process Control Analysis for Monitoring by NIRA, NIRA Seminar, Charlotte, NC, 1986.
28. M. Costello and S. Knepper, Identification of Post-Consumer Carpets by NIRA, *AATCC Textile Applications of NIR Technology Symposium*, Asheville, NC, 1996.
29. W. Jasper and E. Kovacs, *Textile Res. J.*, 64: 444 (1994).
30. K. Beck, Qualitative Applications of NIR — Fiber Identification, *AATCC Textile Applications of NIR Technology Symposium*, Asheville, NC, 1996.
31. J. Rodgers, *AATCC Review*, 2: 27 (2002).
32. A. Elliott, et al., *Brit. J. Appl. Physics*, 5: 337 (1954).
33. J. Rodgers, et al., Process Monitoring of the Moisture and Finish-on-Fiber of Textile Products Using a Portable Near InfraRed Analyzer, *Pittsburgh Conference 2002*, New Orleans, LA, 2002.
34. J. Rodgers and S. Ghosh, NIR Analysis of Textile Fibers and Yarns — from Bench-top to On-Line, *7th International Diffuse Reflectance Conference*, Chambersburg, PA, 1994.
35. J. Rodgers and S. Ghosh, At-Line Application of NIR Spectroscopy for Textiles, *Eastern Analytical Symposium*, Somerset, NJ, 1995.
36. J. Rodgers and S. Ghosh, *Textile Res. J.*, 70: 519 (2000).
37. J. Rodgers and S. Ghosh, *Textile Res. J.*, 71: 135 (2001).
38. S. Ghosh, *J. Textile Institute*, 84: 85 (1993).
39. S. Ghosh and G. Brodmann, *Textile Chem. Color*, 25: 11 (1993).
40. R. Jerry, On-Line NIR Measurement of Pigment Concentration During Color Fiber Extrusion, *AATCC Textile Applications of NIR Technology Symposium*, Asheville, NC, 1996.
41. J. H. Saunders, Polyamides, Fibers, in *Encyclopedia of Textiles, Fibers, and Nonwoven Fabrics* (M. Grayson, ed.), John Wiley and Sons, New York, 1984, p. 347.
42. H. W. Starkweather, Transitions and Relaxations in Nylons, *Nylon Plastics* (M. Kohan, ed.), John Wiley and Sons, New York, 1973, p. 307.
43. E. Scholz, *Karl Fischer Titration*, Springer-Verlag, New York, 1984.
44. J. Mitchell, *Anal. Chim. Acta*, 81: 231 (1976).
45. P. Rotolo, *Cereal Foods World*, 24: 94 (1979).
46. H. Mark, *Chemical Assignments of Near Infrared Absorption Bands*, Customer Applications Group, Technicon, Tarrytown, NY.
47. H. Billica, *Fiber Producer*, April: 21 (1984).
48. R. Crossfield, *Fiber World*, March: 21 (1984).
49. C. King, *Fiber Producer*, October: 17 (1980).
50. Application Note #26, FOSS NIRSystems, Silver Springs, MD.
51. R. Bost, et al., *Textile World*, May: 103 (1982).
52. M. Denton, *Textile Month*, June: 57 (1983).
53. S. Ghosh and J. Rodgers, *Melliand Textilberichte*, 67: 26 (1986).
54. J. Baeder, Operation Techniques on Continuous Heatsetting Machines, Part II: Early Detection of Process Upsets, *Proc. 2nd Symposium on Solving Tufted Textile Problems*, Textile Eng. Dept, Auburn University, Auburn, AL, 1981.
55. K. Gosline, An Overview of the Carpet Yarn Twist Setting Process, *Proc. Carpet Manufacturing Technology Conference*, Clemson University, Clemson, SC, 1984.
56. W. Hotfeld and M. Shepand, *Can. Textile J.*, May: 72 (1977).
57. R. Miller and M. Gibson, Effects of Process Induced Morphology Changes on Dyeing of Nylon6,6 Carpets, *AATCC RA91, International Dyeing Symposium*, Atlanta, GA, 1983.
58. R. Brill, *Makromol. Chem.*, 18/19: 294 (1956).
59. T. Itoh, *Jpn. J. Appl. Physics*, 15: 2295 (1976).
60. H. Starkweather and G. Jones, *J. Polym. Sci., Physics Ed.*, 19: 467 (1981).
61. J. Rodgers, NIRA Potential for the Process Monitoring of Heatset Temperature, *Proc. 10th International NIRA Symposium*, Technicon Technical Center, Tarrytown, NY, 1986.
62. H. Mark and D. Tunnel, *Anal. Chem.*, 57: 1449 (1985).

26 Recent Advances in the Use of Near-IR Spectroscopy in the Petrochemical Industry

Bruce Buchanan

CONTENTS

26.1 Introduction	521
26.2 Fuels	521
26.3 Oxygenates in Fuels	522
26.4 Octane Number	523
26.5 Composition of Fuels	524
26.6 Bitumen Analysis	525
26.7 Environmental/Health Analysis	525
26.8 Conclusion	526
References	526

26.1 INTRODUCTION

Near-infrared (NIR) spectroscopy is attractive for use in the petrochemical industry because it measures the overtone and combination bands predominately of the C—H stretches. This fact alone was responsible for the early use of NIR, before the widespread use of computers, in measuring petrochemicals, where Ellis measured over 40 organic liquids including many petrochemicals [1]. One great advantage in using NIR spectroscopy is that the instrumentation can be separated from volatile solutions using inexpensive fiber optics. Also, the method is relatively rapid and noninvasive. For these reasons, NIR spectroscopy has been used in what many would consider a mature and well-characterized industry to further enhance production and safety. Oil has been gathered (in the case of the early Pennsylvania oil field) or pumped for decades. The ability to refine the components into lubricants and fuels was known in the earlier part of the twentieth century. Recent advances in NIR spectroscopy have made it a valuable tool in the refining process and in the safety of the industry.

26.2 FUELS

Oils have been converted into fuel for centuries. With the first discovery of oil naturally seeping into ponds and rivers in Pennsylvania, the petrochemical industry grew rapidly. With the discovery of

a new cracking method, gasoline production increased even faster [2]. Technological advances in the power and automotive industry have precipitated advancements in fuel requirements, requiring a further evolution of the industry. In many cases, the fuel must burn cleaner or with less noxious by-products. NIR spectroscopy has been applied in the process of developing and characterizing these fuels.

26.3 OXYGENATES IN FUELS

Many states are now requiring that fuels be reformulated depending upon the season. The most common formulation requires the addition of an oxygenate during the winter months for many cities that the Environmental Protection Agency has classified as CO nonattainment areas. By increasing the oxygen level in the fuels usually to at least 2.7% by weight, the levels of CO in combustion is reduced. The most common oxygenates are ethanol (EtOH), methyl *tert*-butyl ether (MTBE), ethyl *tert*-butyl ether (ETBE), and *tert*-amyl methyl ether (TAME). In many cases NIR spectroscopy can aid the analysis or control of the oxygenate levels in the fuels.

The National Institute of Standards and Technology (NIST) has the charge of supplying standard reference materials (SRMs) to industry. NIST currently provides eight reference standards containing the four oxygenates from 2.0 to 2.7% oxygen mass fraction in 20-ml sealed glass ampules. The current method used by NIST to characterize the samples consists of gas chromatography (GC) with flame ionization detection (FID). Because the GC method requires that the ampules be opened, the standards must be batch certified. Of course, losses may occur in the transfer to the vials and in the process of sealing. A spectroscopic technique that is able to scan through the vials would be able to detect outlier samples, which in turn would decrease the uncertainty in the SRMs.

To characterize the gas standards *in situ* [3], a Bruker IFS66 Fourier Transform infrared (FT-IR) spectrometer (Billerica, MA) was used to obtain the NIR spectra. A tungsten light source in conjunction with a silicon-coated CaF₂ beamsplitter and a Ge detector allowed spectra to be obtained from about 12,000 to 5,500 cm⁻¹ (0.83 to 1.8 μm). The spectra were 16 scans co-added to increase the signal to noise ratio. The resolution of the spectra was 4 cm⁻¹. Raman spectra were obtained using a Bruker RFS-100 Fourier Transform Raman (FT-Raman) spectrometer. The source consisted of a 1-W diode-pumped CW Nd-YAG laser (ADLAS DPY 321, Lubeck, Germany) operating at 1.064 nm. Spectra consisted of 16 scans co-added from 500 to 3200 cm⁻¹. Apodization of the interferograms consisted of Blackman–Harris four-point filter and zero-filled by a factor of 4 before transformation.

The gasoline base used in the SRMs was an industry average blend obtained from Phillips Petroleum (Bartlesville, OK). HPLC-grade MTBE was obtained from Aldrich Chemical Co. The ETBE and TAME were obtained from Spectrum Quality Standards with purities of 98.2 and 98.7%, respectively. Anhydrous EtOH was obtained from Warner–Graham Co. (Cooksville, MD). The SRMs were synthesized gravimetrically. Non-SRM standards were mixed in-house. All samples for analysis were contained in 20-ml flame-sealed borosilicate ampules.

Although the ampules were relatively uniform in diameter and thickness, there was some variation that could affect the spectra. An NIR ampule spinner was fabricated to reduce the effects of these differences. An additional preprocessing step included multiplicative scatter correction (MSC). The MSC method reduces the spectra to a “mean” scatter coefficient. This method can be applied to both NIR and Raman spectra.

Partial least squares (PLS) were used to predict the mass fraction of the standards based on mean-centered and MSC-corrected spectra. More than 600 spectra were obtained for use in the study. The models developed were linear with intercepts near zero in all cases. The correlation coefficient (R^2) for the models exceeded .999, and reproduced the calibration sets to within 0.03 oxygen percent mass fraction.

Near-infrared and FT-Raman spectra are capable of quantifying the four common types of oxygenate gas additives. There is sufficient accuracy to detect potential outliers, thus tightening the confidence intervals for the SRMs.

Another study compares the use of FT-Raman, FT-IR, and dispersive NIR analysis for the analysis of oxygenates, in this case MTBE [4]. The study used 69 commercial gasoline samples (Ashland Refinery, Cattlesburg, KY) with MTBE added in the range 0.2 to 3.262%. The FT-Raman instrument was a Nicolet 950 FT-Raman spectrometer using sNd:YLF laser (1064 nm) with an EGe cooled detector. The spectra were the average of 200 scans at an 8 cm^{-1} resolution. The FT-IR instrument was a Midac Fox with a Deuterated Triglycine Sulfate (DTGS) detector and a SiC source. The spectra were the average of 40 scans at 4 cm^{-1} resolution. The sampling system consisted of a ZnSe attenuated total reflectance (ATR) sample cell. The NIR instrument was a NIRSystems Online 6500 dispersive instrument (Foss, Silver Spring, MD) using a PbS detector. The spectra were the average of 32 scans acquired at 2 nm resolution. The sampling system consisted of a fiber optic bundle with a transfectance stainless steel probe.

In all three cases a PLS model was developed on mean-centered data. The number of optimal factors was determined using leave-one-out cross-validation. The FT-Raman model used the regions of 3170–2559, 863–848, 736–697, and 532–497 cm^{-1} . The FT-IR model used the regions of 3133–2660, 1412–1308, 1279–1138, 1117–1065, and 872–831 cm^{-1} . The NIR model used 2068–1812, 1646–1620, 1548–1328, and 1286–1114 nm. The spectrum of MTBE was used to select the optimal regions. In addition for the NIR method, all regions with an absorbance greater than 2 were eliminated. For all three methods five factors were found to be optimal. The article summary lists the SEP for the methods as 0.155, 0.143, and 0.131 wt.% for FT-Raman, FT-IR, and NIR, respectively. While these values are nowhere in the body of the article, it does show in Table II of the referenced paper values of 0.169 (five factors), 0.143 (four factors), and 0.141 (six factors) for FT-Raman, FT-IR and NIR. While these values are denoted as SEV, inferred to be the standard error of validation, the text refers to them as SEP. The higher SEPs in this study, as compared to the NIST study, is understandable as the range here is 0.2 to 3.262 wt%, and the NIST study used a much narrower range. All three methods gave similar results but both Raman and NIR have an advantage over FT-IR in that optical fibers can be used to isolate the equipment from the sample. A decided advantage is in gasoline analysis!

26.4 OCTANE NUMBER

The current, established method of determining octane number consists of two tests according to the American Society for Testing and Materials (ASTM). The Research Octane (RON) method [5] consists of an ASTM-certified engine with set intake temperatures and speeds. The Motor Octane (MON) method [6] uses the same ASTM-certified engine but is set at a different intake temperature and speed. The value reported at the pump is an average of the two values. The engine, costing over U.S. \$250,000, is expensive and there is an additional expense in the maintenance. Because the octane value is related to the proportion of straight-chain hydrocarbons to branched and aromatic hydrocarbons, NIR analysis is feasible. This is, of course, because the NIR region consists predominantly of the overtone and combination bands of the C–H stretch. In addition to increased safety and time savings, the use of NIR spectroscopy to determine the octane value also generates a potentially large monetary savings.

A standard error of calibration (SEC) of 0.18 MON and an SEP of 0.22 MON was obtained using a tunable acousto optic filter based instrument [7]. The spectral range measured was 900 to 1700 nm. Multiple linear regression (MLR) analysis was used to correlate the motor value with the NIR value. The paper did not specify how many samples were used in the analysis or verification set or if any preprocessing steps were used.

Ridge regression is used to determine the RON number using NIR spectra [8]. Fifty-eight different unleaded gasoline samples were prepared randomly by blending nine different stocks. The samples ranged in RON values between 90.5 and 98.4. The study compared two methods, a four-factor MLR method with a four-factor ridge regression. While the two methods were comparable, the ridge regression method was more robust. Both methods give rise to SEPs in the range of 0.25 to 0.3. The advantage of both chemometric methods is that the use of four wavelengths allows the use of diode instruments.

Naturally, a great advantage would arise if the octane measurement could be moved on-line. In the overview written by Zetter and Politzer, the advantages of on-line measurements are laid out [9]. In addition to the advantages discussed above, the use of on-line analysis reduces the “Octane Giveaway.” By narrowing the spread of octane values around the target, lesser quantities of higher octane fuels need to be blended. In an NIR project by British Petroleum lasting over 3 years, 12,000 NIR spectra and the associated RON and MON values were obtained using FT-NIR technology [10]. The data showed a good correlation between the octane values even though the data were obtained from multiple refineries.

26.5 COMPOSITION OF FUELS

NIR spectroscopy has been used to monitor the percent aromatics, percent saturates, and freezing points in JP-4 aviation fuels [11]. The control of aromatics in jet fuels is critical because fuels with lower aromatic content are more efficient and produce less exhaust color, and the fuels are less corrosive. Being less visible and preserving o-rings, gaskets, and lines provide a great advantage to fighter jets! Typically, the ideal aromatic content is less than 25%, which is well within the range of NIR spectroscopy. Spectra of jet fuel samples were obtained in the “short wavelength” region (850 to 1100 nm) using an HP-8452 diode array instrument with a 2.00 cm sample cuvette. The “long wavelength” (1100 to 1500 nm) spectra were obtained using a NIRSystems (now Foss, Silver Spring, MD) model 6250 scanning spectrophotometer equipped with a 1.00 cm sample cuvette. Using a second-derivative preprocessing and stepwise multiple linear regression (SMLR) the authors were able to predict aromatics ($\text{SEP} = 1.70$, $R^2 = .86$), saturates ($\text{SEP} = 1.68$, $R^2 = .85$) and freezing point ($\text{SEP} = 3.92$, $R^2 = .85$) using the short wavelength range on a sample set of 36 fuels. Using the long wavelength region, the authors were able to predict aromatics ($\text{SEP} = 1.82$, $R^2 = .86$), saturates ($\text{SEP} = 1.88$, $R^2 = .85$) and freezing point ($\text{SEP} = 4.94$, $R^2 = .76$). The sample set ranged from 5.68 to 25.74% aromatics, 70.66 to 90.55 vol% saturates, and freezing points of -70.95 to -38.33°C .

In another study, NIR spectroscopy was used to analyze a set of 359 gasoline samples for RON, MON, aromatics, olefins, saturates, MTBE, EtOH, oxygen, benzene, and sulfur [12]. The samples were obtained from a diverse set of retail locations (23). Both SMLR and PLS were used to correlate the values to the NIR spectra. The data set was divided into a calibration set of 269 and a verification set of 90 samples. The results are reproduced in Table 26.1. Most values were predicted reasonably. The exception was sulfur, which is not unexpected as sulfur is not an NIR active species.

A carefully designed NIR analysis system was used to optimize the blending of gasoline [13]. A NIR analyzer is used at a St. Paul, MN refinery with an on-stream time greater than 99%. The analyzer was designed to contribute an error of less than 0.02 octane number to the final result. A three-year study shows an SEP using the NIR method of 0.28 for RON, 0.24 for MON, and 0.21 for road octane. The ASTM SEP for the same values were 0.41, 0.61, and 0.36, respectively. The study also discusses the various design parameters that are involved in designing an NIR analysis system for gasoline blending.

Asphaltenes molecules are considered to be dispersed in oil by the more polar molecules [14]. Under certain conditions, the asphaltene molecules can deposit out causing production or transportation problems. NIR spectroscopy using a high pressure coupled with PCA analysis was used to monitor the onset of asphaltene aggregation. A Brimrose Acoustic Optic Tunable Filter instrument was used to obtain the spectra. An Equitech high-pressure (maximum of 325 bar) sample cell with a 1.5 mm pathlength was used to sample the solution. The first two PCA vectors describe 99.7% of the spectral information and show that the asphaltene aggregation occurs at about 180 bar for a crude oil sample. NIR spectra were used further to show that the asphaltene aggregation is reversible by increasing the pressure beyond the aggregation point, although it did take 72 h to equilibrate.

TABLE 26.1
Gasoline Properties Using NIR Spectroscopy

Property	SMLR model		Number factors	PLS		ASTM 1σ
	R^2	SEP		R^2	SEP	
RON	.964	0.53	8	.984	0.35	0.30
MON	.959	0.43	6	.972	0.37	0.45
Aromatics, %	.917	1.77	5	.921	1.17	1.7
Olefins, %	.853	1.96	8	.933	1.41	2.6
Saturates, %	.869	2.18	8	.871	2.26	2.7
MTBE	.946	0.97	7	.991	0.44	0.79
EtOH	.999	0.11	5	.985	0.43	0.93
Oxygen	.980	0.15	7	.989	0.11	—
Benzene	.623	0.29	7	.952	0.14	0.41
Sulfur	.363	0.012	11	.546	0.013	0.005

Note: SMLR model consisted of four wavelengths except for the EtOH, which consisted of two wavelengths. The PLS model was developed with a second-derivative preprocessing.

26.6 BITUMEN ANALYSIS

Oxidized or degraded oil sands show poor processing characteristics [15]. These are known as bad ores, problem ores, type X ores, or ores with a high misery factor (our favorite)! These high misery factor ores then need to be processed with extensive process aids at an additional cost. Microscopic analysis of the “froths” can be used to distinguish the oxidation state of the bitumen, which then correlates to the grade of the oil sand. As the bitumen degrades, the $-\text{CH}_2$ peak 1750 nm decreases when the $-\text{OH}$ peak at 1950 nm increases, showing the basis of using spectroscopy. The NIR spectra obtained on the froths of a set of known sand types were correlated to the microscopic results with an R^2 of 0.8883. The advantage of NIR spectroscopy is that it does not require a froth separation step. By typing the ores, the use of process aids can be reduced giving rise to significant savings.

26.7 ENVIRONMENTAL/HEALTH ANALYSIS

Typically, NIR spectroscopy is not used for environmental analysis as environmental contaminants are measured in the parts per million (ppm) or less. However, there have been reported studies on the use of NIR spectroscopy in the contamination of soils with oil and diesel fuels.

In a study done on sandy loams, it was shown that the detection limit of oil at 1 sigma was 0.17 wt% [16]. The current method uses solvent extraction with gas chromatography. While the NIR method would be unable to match the PPM level of the GC method, it could be used as an inexpensive screening method to reduce the number of GC analysis that must be performed. To verify the feasibility, sample set was synthesized from a sand sample obtained in an open area at the Sandia National Laboratory near Albuquerque, New Mexico (obtaining a sand sample near Sandia National Laboratory is a relatively simple endeavor)! The sample was sieved, dried, and equilibrated to ambient laboratory atmosphere. Oil was added to the sample in known amounts and the spectra obtained on a Nicolet 800 Fourier transform spectrometer with a third-party (SpectraTech) reflectance attachment and an InSb detector. An SEP of 0.17% was obtained using a two-factor PLS model on a set of 17 samples. The SEP was calculated using the typical leave-one-out cross-validation. In addition, work was performed on potting soil and garden soil. As these two types of soils have a

significantly higher organic component, there was some concern that this material would interfere with the analysis. Developing a set of synthetic spectra with added components of the other soil types shows that the SEP remains at 0.17%; however, the authors do warn that to develop a calibration that could adequately predict oil contaminants these types of soils would probably have to be in the calibration equation.

In a related study, diesel contamination of soils in the Manitoba region is predicted using NIR spectroscopy compared to solid-phase microextraction GC [17]. The set of 26 samples were divided into two sets of 13 spectra each. SMLR was used to predict the diesel value given as total petroleum hydrocarbon (TPH). The model developed on set A was used to predict set B and then reversed. The statistics were then pooled, with an R^2 of 0.71 and an SEP of 0.77 mg/g. While this SEP is relatively high (with a sample range of 0.001 to 6.26 mg/g) this is most probably attributable to the high variation in the reference method. Because there is varying weathering in the samples, the diesel components differ in a large degree. Also, the reference method has a high variation; a comparison of the results of 36 laboratories gave an order of magnitude error when measured on an identical sample. Under these circumstances the correlation and SEP show promise for the method.

Lubricants are a significant fraction of the petrochemical industry. Unfortunately, lubricants can contain a significant fraction of cancer-causing compounds, most notably polynuclear aromatic (PNA) compounds. In fact the European Union (EU) requires that lubricants that contain more than 3% PNA (w/w) must be analyzed using the DMSO extract method. Given the relatively high percentage, NIR should be a useful technique in the analysis of lubricants. An NIR method was developed using an Acousto-Optical Tunable Filter (AOTF) instrument coupled with principal component regression (PCR) to measure the aromatic components in lubricating oils [18]. The method used a Luminar 2040 (Brimrose) with an InGaAs detector and a 1 cm pathlength cell. The spectra of 44 lubricant base oil samples were the average of 300 scans between 1300 and 2500 nm. The wavelengths 1300–1354, 1394–1462, and 1780–2246 based on the PC vectors were actually used. The first 5 PC explained 99.3% of the information in the spectra. A comparison of preprocessing methods (derivatives, normalization, or MSC) showed that the normalization ($R = .995$, root mean standard error of prediction (RMSEP) of 0.4%) or the MSC ($R = .994$, RMSEP of 0.5%) preprocessing gave rise to the best predictive models using 6 PC vectors. Using NIR spectroscopy is a quick and inexpensive method to classify lubricant oil samples.

26.8 CONCLUSION

The petrochemical industry is a fairly “mature” industry. Oils, kerosene, and gasoline have been processed since the 1800s. Many process improvements and much process knowledge have been accrued from the start of the industry. Even so, process improvements and safety enhancements are still possible and NIR spectroscopy is one tool that can be an aid. With the improvements in the instrumentation and computers, the ability to use NIR spectroscopy increases even more. Several papers, although by no means all inclusive, have been presented to demonstrate the wide range of applications in the petrochemical industry for which NIR spectroscopy is a useful tool.

REFERENCES

1. Ellis, J. W., *Phys. Rev.*, 22: 200 (1923).
2. Burton, W. M., US Patent 1,049,667, 1912.
3. Choquette, S. J., Chesler, S. N., Duewer, D. L., Wang, S., and O'Haver, T. C., *Anal. Chem.*, 68: 3525 (1996).
4. Cooper, J. B., Wise, K. L., Welch, W. T., Bedsoe, R. R., and Sumner, M. B., *Appl. Spectrosc.*, 50: 917 (1996).
5. ASTM D2699, Standard Test Method for Research Octane Number of Spark-Ignition Engine Fuel, ASTM International.

6. ASTM D2700, Standard Test Method for Motor Octane Number of Spark-Ignition Engine Fuel, ASTM International.
7. Muller, D., *Am. Chem. Soc., Div. Fuel*, 43: 287 (1988).
8. Chung, H., Lee, H., and Jun, C.-H., *Bull. Korean Chem. Soc.*, 22: 37 (2001).
9. Zetter, M. S. and Politzer, B. A., *Hydrocarbon Process.*, 72: 103 (1993).
10. Brown, D., *Petroleum Rev.*, 569: 281 (1994).
11. Lysaght, M. J., Kelly, J. J., and Callis, J. B., *Fuel*, 72: 623 (1993).
12. Swarin, S. J. and Drumm, C., *Spectroscopy*, 7: 42 (1992).
13. Welch, W. T., Bain, M. L., Maggard, S. M., and May, J. M., *Oil Gas J.*, 92: 48 (1994).
14. Aske, N., Kallevik, H., Johnsen, E. E., and Sjöblom, J., *Energy & Fuels*, 16: 1287 (2002).
15. Mikula, R. J., Munoz, V. A., Wang, N. et al., *J. Canad. Petroleum Technol.*, 42: 50 (2003).
16. Stallard, B. R., Garcia, M. J., and Kaushik, S., *Appl. Spectrosc.*, 50: 334 (1996).
17. Malley, D. F., Hunter, K. N., and Webster, G. R. B., *J. Soil Contam.*, 8: 481 (1999).
18. Lima, F. S. G., Araújo, M. A. S., and Borges, L. E. P., *Tribol. Int.*, 36: 691 (2003).

27 NIR Analysis of Polymers

Cynthia Kradjel and Kathryn A. Lee

CONTENTS

27.1	Introduction	530
27.2	Comparison of Near-Infrared vs. Mid-Infrared for Polymer Analysis	530
27.2.1	Overtones	531
27.2.2	Combination Bands	532
27.2.3	Intact Sampling	532
27.2.4	Identity and Quality Testing	533
27.2.5	Process Analysis	534
27.3	Polymer Applications of Near-Infrared Analysis	535
27.3.1	Polyethylene and Polypropylene	535
27.3.2	Polyvinylchloride	536
27.3.3	Polyvinyl Alcohol	537
27.3.4	Polyvinyl Acetals	539
27.3.5	Polystyrene	539
27.3.6	Acrylic and Methacrylic Polymers	541
27.3.6.1	Ester-Based	541
27.3.6.2	Amide-Based	542
27.3.7	Polyamides	543
27.3.8	Cellulose Ester	544
27.3.9	Substitution on Starch	544
27.3.10	Alcohols	545
27.3.10.1	Fatty Alcohols (Higher Aliphatic Alcohols)	545
27.3.10.2	Synthetic Alcohols	547
27.3.10.3	Polyhydric Alcohols	547
27.3.10.4	Esterified Alcohols	548
27.3.10.5	Ethoxylates	548
27.3.10.6	Alkoxy Alcohols	548
27.3.10.7	Sulfonated Alcohols	550
27.3.10.8	Method for Analyzing a Wide Variety of Alcohols	551
27.3.11	Polyurethanes	551
27.3.12	Polyesters	551
27.3.12.1	Alkyds	551
27.3.12.2	Unsaturated Polyesters	552
27.3.12.3	Saturated Aliphatic Polyesters	552
27.3.12.4	Aromatic Polyesters	552
27.3.13	Epoxy Resins	553
27.3.14	Phenolic Resins	561
27.3.15	Silicones	561
27.3.16	Formaldehyde	562
27.4	Conclusions	563
	References	563

27.1 INTRODUCTION

Although Leo Baekeland produced the first true synthetic plastic in 1907, the scientific community as a whole did not accept macromolecular theory until the late 1920s. Staudinger [1–4] first proposed the theory of large molecular weight compounds in 1920. He synthesized many macromolecules and then tested viscosity, melting points, and solubility. A debate raged between Staudinger and other proponents of *macromolecular theory* vs. proponents of the *association theory*. The association theorists rejected the possibility of high molecular weight compounds and did not believe that these atoms were linked by covalent bonds.

Instrumental analysis provided the proof that ended the controversy. Dr Herman Mark presented his x-ray crystallography studies in Germany, in 1926, concluding that “cellulose has covalent bonds” [5]. Detection of both crystalline and noncrystalline regions in cellulose led to the discovery that polymers do not have a uniform molecular weight but a range of molecular weights that can be statistically averaged.

Today polymers scientists rely on many different methods of instrumental analysis to develop, test, and classify macromolecules. Infrared (IR) spectroscopy, one of the most important analytical tools in research as well as in manufacturing, is used to test the quality of incoming monomers, to control processes, to control quality control, and to confirm quality of intermediates and final products. IR is even used to troubleshoot processing problems and to analyze the competitor’s products. IR spectroscopy has value in structure elucidation and quantitative analysis of polymers. Hausdorff [6] and Kagarese and Weinberger [7] were the first to publish collections of IR spectra of polymers.

The near-infrared (NIR) region of the electromagnetic spectrum was discovered by William Herschel [8] in 1800. However, it was not until the 1950s that suitable sources and sensitive lead sulfide detectors [9] were available for NIR instrumentation. In 1954, engineers at DuPont developed an NIR process analyzer using the newly discovered lead sulfide detector and an Ebert monochromator for dispersion. The same group developed a “workhorse” NIR process analyzer using the lead sulfide detectors and optical interference filters [10].

Early work in the field of NIR analysis was published by Robert F. Goddu [11] at Hercules Powder Company, Wilbur Kaye [12] at Tennessee Eastman Company, and Owen H. Wheeler [13] at the University of Puerto Rico. These reviews are extensive and include detailed band assignment of organic functional groups in the NIR region. The authors prepared the samples by dissolving them in solvents such as CCl₄ prior to analysis. Quantitative analysis was performed using simple peak ratio calculations. Glatt and Ellis [14] published studies on linear polymers. Miller and Willis [15] did early work on quantitative analysis of synthetic polymers.

Elbert Crandall [16] at Kansas State College found the NIR region valuable for studying polyamic acids as well as step reaction and addition polymers. Crandall melted the polymers and pressed them between glass plates to form a transparent film. Most NIR analyses of polymer today are performed directly on the material without sample pretreatment as a result of advances in instrument design, sample handling devices, and the development of computer-generated chemometric data analysis. Sophisticated algorithms using factor analysis enable users to “decompose” the data and determine information about the parameter of interest from spectra that contain information about the entire matrix as well.

27.2 COMPARISON OF NEAR-INFRARED VS. MID-INFRARED FOR POLYMER ANALYSIS

The NIR region complements the mid-IR region [17]. Siesler [18] published studies comparing the IR and NIR region. Due to the different types of absorption bands and their resulting intensities, some polymer applications are better suited to the NIR region of the spectrum, while other applications are better suited to the mid-IR region. The bands in the NIR (0.8 to 2.5 μm) are overtones and

combinations of the fundamental absorbances found in the classical mid-IR region. These bands are mainly due to hydrogenic stretches of C—H, N—H, and O—H bonds. Because organic polymers are composed primarily of carbon, hydrogen, nitrogen, and oxygen atoms, the NIR spectra of organic polymers feature sharp, strong absorbance bands when compared to NIR spectra of natural products.

27.2.1 OVERTONES

Overtone or harmonics are multiples of fundamental mid-IR absorption bands. The approximate location of the first and second overtone can be calculated by dividing the frequency of the fundamental by a factor of two and by a factor of three respectively. (The actual frequency of the harmonics are shifted about several percentage points away from the exact multiple value due to the anharmonicity constant in Schroedinger's equation.)

The relative intensity of the absorbance band of the first overtone of a fundamental is approximately one order of magnitude less than the relative intensity of the fundamental absorption band. The relative intensity of the absorbance band of the second overtone is approximately two orders of magnitude less than the relative intensity of the fundamental absorption band and so on.

The diminished intensity of overtone bands is analytically useful for spectroscopic applications involving analysis of samples containing strong mid-IR absorbers; nonhomogeneous samples that require averaging greater surface area to improve results; and samples containing inorganic fillers.

Samples containing strong mid-IR absorbers, such as aqueous solutions, can be analyzed directly. Polymer latexes with up to 99% water content have been analyzed for percent solids and copolymer ratio. The overtone bands of water are still very strong, but are confined to the region from 1400 to 1500 nm leaving the (C—H)_n and (N—H)_n overtone region (1500 to 1900 nm) relatively unaffected by the large concentration of water (Figure 27.1). Sample cells with variable pathlengths can be used to increase the sensitivity of the organic constituents.

Packing materials, including laminates and other types of multilayered films, can be analyzed intact because the NIR light goes through all layers. Absence of any component in a laminate of polypropylene, polyethylene, and nylon can be detected by NIR [19].

Pellets or molded products can be analyzed “as is” [20]. When the absorption band is off-scale in the primary overtone region (1300 to 1900 nm) then the secondary overtone region (700 to 1300 nm) can be utilized.

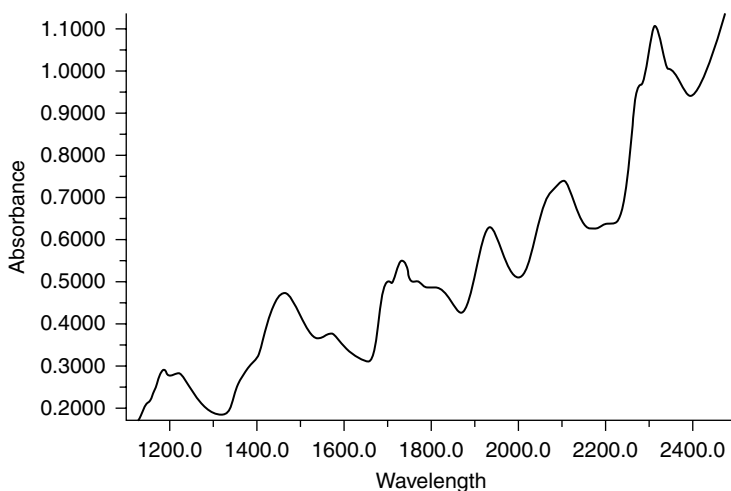


FIGURE 27.1 Latexes can be analyzed “as is” for copolymer ratio and solids content. In this scan of a typical latex, the C—H and N—H bands of the polymer can be seen between 1500 and 1900 nm and above 2200 nm. The O—H of water can be seen between 1400 and 1500 nm and 1900 and 2000 nm.

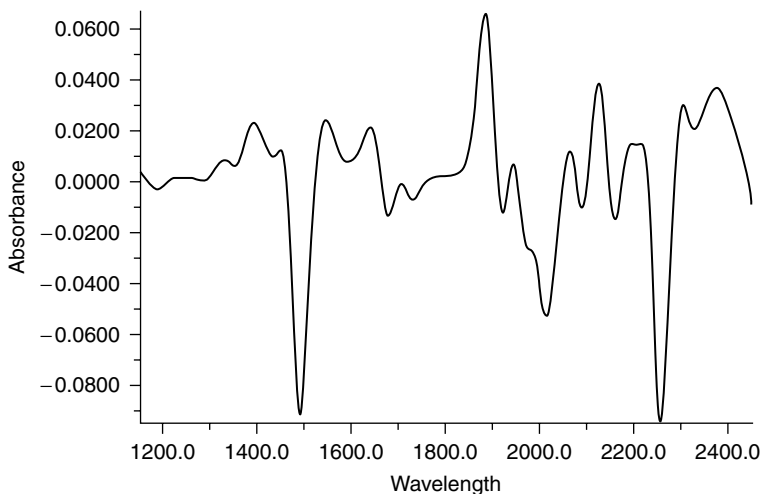


FIGURE 27.2 Second-derivative scan of a melanine/formaldehyde glass-filled composite. NIR can be used to analyze for percent binder and degree of cure without having to destroy the sample.

Reinforced thermoplastics or composites are often nonhomogeneous and require much averaging to get a representative result. Overtone bands of nonhydrogenic bonds of inorganic compounds are very weak and do not absorb appreciably in this region. Therefore, NIR is useful for the nondestructive determination of organic compounds in the presence of inorganic fillers, such as percent binder and degree of cure in composites (Figure 27.2).

Another advantage of the lesser intensity of inorganic bands is the lack of absorbance of glass, enabling its use in sample devices. Samples can be pressed between disposable microscope slides and scanned, or scanned through glass vials. Fiber-optic sampling devices, manufactured from glass, are widely used in NIR analysis for easy sampling of raw materials or for *in situ* measurement of process streams. Note, however, if an application requires measurement in the region where Si—OH absorbs (2100 nm), glass is not suitable as a sampling device. The glass will retain varying amount of water, resulting in a variable peak at 2100 nm due to Si—OH absorbance.

27.2.2 COMBINATION BANDS

Combination bands arise from sums and differences of fundamental stretching and bending vibrations. The combination bands of molecules are dependent on molecular symmetry and thus are unique [13]. As a result, absorbance due to sums and differences of molecular motion can be used to resolve closely absorbing species. For example, consider a sample matrix containing the O—H bands of alcohol, carboxylic acid, and water. The overtones of all three O—H bands fall within tens of nanometers away from each other whereas the combination bands fall approximately 100 nm apart (Figure 27.3).

Another example of the uniqueness of combination bands is the resolution of the O—H and N—H bands as seen when monitoring epoxide curing with NIR spectroscopy. The O—H and N—H fundamental bands often overlap in the mid-IR whereas they are distinguishable in the NIR region. Further discussion of NIR analysis of epoxy curing reactions is covered later in this chapter.

27.2.3 INTACT SAMPLING

The greatest advantage of NIR spectroscopy is the ease of sample handling. Reflectance, transmission, or transmittance can be used. Classical spectroscopy requires physical separation of the

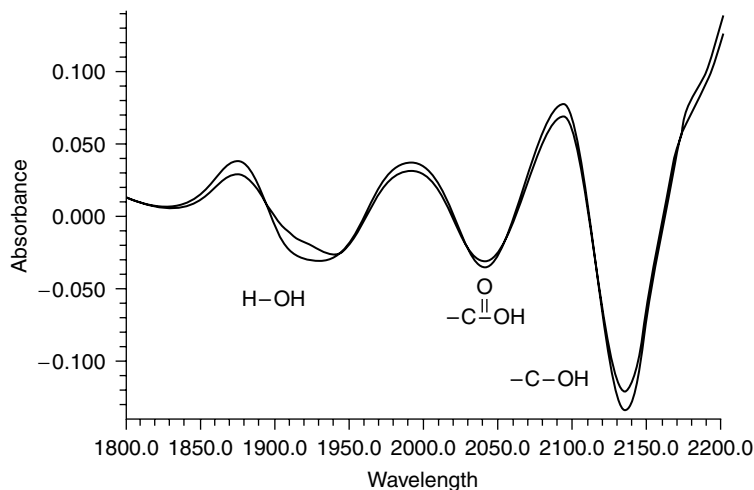


FIGURE 27.3 Second-derivative representation of two polyester resins in the OH combination band region. The O—H bands of water, alcohol, and carboxylic acids are well resolved between 1900 and 2200 nm.

constituent of interest from the matrix, usually by dissolution in a solvent. Conversely, NIR spectroscopy when combined with chemometric principles does not require physical separation of the analyte from the matrix. Absorption bands due to the constituent of interest are “mathematically” separated from the absorption band of the matrix.

Nondestructive analysis of the sample also makes NIR spectroscopy unique in that physical and chemical properties can be measured simultaneously. $\text{Log}(1/R)$ spectra have been correlated to a variety of physical properties such as particle size [21], viscosity of polymers, degree of dispersion in paint system [22], and “heat history” of nylon [23], a property related to the crystalline/amorphous ratio of nylon.

27.2.4 IDENTITY AND QUALITY TESTING

Advances in factor analysis have lead to the increased usage of qualitative analysis for polymer analysis. A teaching set of known materials is used to determine acceptable radii and residual criteria using either library or cluster modeling techniques [24]. Identification and verification of unknown substances is then achieved by comparing the radii and residual criteria of the new material to the criteria established for the reference materials in the modeling step.

Howell and Davis [25] demonstrated the use of NIR for discriminating between different types of polymer materials.

NIR techniques are also used for identity confirmation of fabricated products such as cap liners that are used in drinking bottles. Eilert [26] describes a method with ten liner types including polyethylene, polypropylene, polyester, and a polyethylene/polypropylene blend. Cluster models were successfully established to identify the material used in unknown cap liners. This model also included several materials that differed only in quality or “grade.” With the proper teaching set, the NIR qualitative method can be established to separate between different quality or grades of material.

NIR qualitative methods have also been applied in blending stages of polypropylene-powdered polymer blends for rapid discrimination of “in spec” and “out of spec” material [26]. The teaching set used samples that were correctly proportioned as well as samples that had either 20% too little or too much of one the five additives. Concentration targets ranged from about 2% to about 13% of additive. The model was able to correctly classify unknowns as either “in spec” or “out of spec.”

Identification of plastics waste by NIR has been studied by several groups. The Fraunhofer Institute [27] has developed a method based on an analyzer using acousto-optical tunable filter (AOTF) techniques, operating between 1000 to 1800 nm. The identification rate exceeds 10 parts per second with part sizes of 60 mm and distances between the parts of 50 mm for household waste. A method for recognition of PE, PP, PS, polyvinylchloride (PVC), and polyethylene terephthalate (PET) materials in household waste has been developed as well as a method for recognition of PC, acrylobutadiene-styrene (ABS), PMMA, PUR, and PBT in technical plastics. Camacho et al. compared methods for the determination of the composition of binary blends of recycled polypropylene/polyethylene (PP/HDPE) and recycled poly(acryl-butadiene-styrene) and polypropylene (ABS/PP). They found on-line diffuse reflectance (DR) near-infrared spectroscopy (NIRS) offered high sensitivity, ease of operation, and the option of multivariate data analysis. In comparison, differential scanning calorimetry (DSC) and mid-IR, which are commonly used for this purpose required certain sample preparation and were slower. The root mean square error of prediction (RMSEP) was 0.21% wt. for 0 to 15% wt. of PP in high-density polyethylene (HDPE) and a RMSEP equal to 0.91% wt. for 0 to 100%. For blends of PP/ABS the RMSEP was 0.74% wt. for 0 to 100% and 0.32% wt. for 0 to 15% wt. PP [28].

Various methods have been developed to separate plastics from nonplastics. van de Broek et al. [29] combined partial least squares (PLS) and multiple linear regression (MLR) data analysis [30] with NIR imaging to achieve the separation. Wienke et al. [31] studied the use of NIR imaging with six-wavelength analysis. They used reference and dark current spectral images to correct the raw spectral images for normal fluctuations in background due to instrumentation conditions [32]. Additional investigation was conducted using FuzzyARTMAP classification models [33,34].

27.2.5 PROCESS ANALYSIS

Another major advantage that the NIR region offers over other techniques is the increased pathlength of anywhere from 0.1 mm to several centimeters. The larger pathlengths reduce clogging and other problems related to particulate content of viscous materials.

“Processability” of polymers can be analyzed using discriminant analysis. Discriminant analysis groups samples in multidimensional space based on the $\log(1/R)$ data of their NIR spectra [35]. Mahalanobis distance, a multidimensional distance measurement, is used to classify unknowns in relation to the groups. In a process situation, the analyzer is calibrated with representative samples of a “rejected” batch, a “borderline” batch, and an “in-spec” batch. New batches can be analyzed as they are charged and processing problems can be detected while the problems are still correctable. These developments are of great interest to manufacturers of thermosetting resins and other irreversible polymeric processes.

Polymer reaction monitoring has been studied by Siesler [36]. Optical-fiber remote NIRS has been used to monitor the composition of methyl methacrylate (MMA) during the polymerization process, the copolymer reaction of styrene/maleic anhydride with 6-aminohexanoic acid, and the structural changes during the formation of PET film.

Several authors describe the use of NIR in severe conditions during polymer melt formation, including Williams and McPeters [37], Brimmer et al. [38], and Lammers et al. [39]. Van der Elst [40] describes the implementation of online acousto-optical-tunable-scanning (AOTS) technology for control of a polymerization process in a batch reactor. He describes the de-bottlenecking and estimated reduction in average cycle time by greater than 10% achieved by online determination of solvent/monomer ratios. The analyzer, an InfraPrime, is interlocked with catalyst and monomer inlet valves, with a response time of less than 15 s. Dual optical fibers are used, one for reference and one for sample beam, to eliminate optical window fouling problems. The system operates with minimum requirements of 15 barg/150°C.

An in-line sensor using NIR spectroscopy was developed for monitoring carbon dioxide (CO₂) concentration, which was dissolved into molten polymers in extrusion foaming processes. The sensor

infers the CO₂ concentration from the measured NIR absorption spectra. The spectra were acquired on-line through the probes equipped at the die of the extruder, where foams were created by dissolving CO₂ gas into molten polymer at the middle of the extruder and releasing pressure at the exit of the die. The experiments were conducted on polypropylene (PP), low-density polyethylene (LDPE), and polybutylene succinate (PBS), at different flow rates and at different operating temperatures. The use of wavelet transform to analyze the NIR spectra allowed elimination of the influence of the operating conditions such as operating temperature and flow rate on CO₂ measurement [41].

Because of the speed of data acquisition of NIR instrumentation with nonmoving parts, opportunities arise to study reaction rates and mechanisms. One example includes the crosslinking reaction of liquid carboxylated poly(acrylonitrile-Co-butadiene) or nitrile rubber (NBR) with dicumyl peroxide (DCPO). This reaction was studied by electron spin resonance (ESR) spectroscopy and the crosslinking reaction was followed *in situ* in dioxane by monitoring of the disappearance of the pendant vinyl double bond with Fourier transform NIR (FT-NIR). The overall activation energy and rate equation of the reaction were able to be determined and provided insight into the reaction mechanism [42].

27.3 POLYMER APPLICATIONS OF NEAR-INFRARED ANALYSIS

27.3.1 POLYETHYLENE AND POLYPROPYLENE

NIR can be used in the feed preparation step to monitor the purity of incoming materials and to test for water content of raw materials. In the addition reaction step, the disappearance of the double bonds can be monitored using NIR. Terminal double bonds have a combination band at about 2110 nm and a first overtone band at 1620 nm. In the recovery phase, the residual solvent can be determined by NIR.

NIR is not well suited for analysis of the backbone structure of the final product. One can measure the ratio of CH₃, CH₂, and CH bonds [43], but their specific location on the molecular backbone cannot be readily determined by this technique.

Macho et al. applied multivariate control techniques to detect outliers in the determination of ethylene in impact polypropylene samples by NIR spectroscopy and multivariate calibration PLS. An algorithm was used that identifies those spectral variables responsible for the outlier behavior so that the source of the behavior could be identified. The outliers in the prediction step may be due to errors associated with the prediction of analyte concentrations in samples which are beyond the concentration range, changes in the matrix composition, and instrumental errors [44]. Plots of the Hotelling T² and Q (sum of squares of residuals), and leverage statistics can be used for analyzing the calibration quality. A high value in the T² plot but not in the Q plot may be a warning that the calibration model is being extrapolated, while a high value in Q and not T² may be because a variability not considered in the PCA model is being incorporated into the signal recorded.

NIR has been used to distinguish between HDPE, LDPE, PP, PET, and PS and blends using discriminant analysis. Samples were analyzed directly in the pellet form using the InfraProver, a FT-NIR system. Furukawa et al. [45] used NIR DR and FT Raman to measure 12 kinds of block and random PP copolymers with different ethylene content in pellets and powder states to predict the ethylene content and to deepen the understanding of the spectra of PP. Band assignments were proposed based on the calculation of the second derivatives of the original spectra, analysis of loadings and regression coefficient plots of PCA and PCR models, and comparison of the spectra of PP with those of linear low-density polyethylene (LLDPE) with short branches. MSC treatment improved the SEP. The NIR chemometric evaluation of the data gave better results than those derived from the Raman spectra and chemometric analysis.

A quantitative method has been developed for ethylene-propylene copolymers [46–48]. Combination bands 2275, 2312, and 2466 nm were used to determine monomer ratios. Miller et al. [49] determined the composition of high-density/LDPE blend films.

Methods have been developed to determine melt flow index of polypropylene beads. Samples were averaged using a rotating cup sample drawer. The predictability of the calibration was enhanced when the learning set was split into two subclasses, “low melt flow index” (0 to 1) and “high melt flow index” (>1). Wavelengths selected for low melt flow index were 1688, 2204, and 2220 nm; r^2 was 0.9998, and standard error of estimate (SEE) was 0.005. The high melt flow index method used wavelengths 1192, 1692, and 2140 nm with an r^2 of 0.9996 and SEE of 0.48 [26]. NIR can be used on-line to monitor the density of molten LLDPE. The calibration models used PCR. The standard error of prediction (SEP) for the on-line monitoring of the density of LLDPE at the plant was $\pm 2.1 \text{ mg/cm}^3$ (range: 0.91 to 0.95 g/cm^3) [50].

Molecular weight of recycled HDPE as measured by size-exclusion chromatography can be correlated with diffuse reflectance near- and mid-infrared spectroscopy (NIR/MIR). The RMSEP obtained with MIR data was 360 and with NIR data was 470. The degree of crystallinity determined by DSC can also be correlated to the spectra. The prediction ability of this model was poor in the 1100 to 1900 nm region, however the 2000 to 2500 nm region yielded better results. Four PC factors explained the most of the variance in the spectra and the RMSEP was 0.4% wt. [51].

NIR is used as a quality control tool for quantifying the additives used in the finishing stage. Quantitative analysis of ultraviolet (UV) stabilizer in polypropylene by NIR is one such example. The combination bands of the methylenic stretches in polypropylene occur between 2000 and 2400 nm, the first overtones occur around 1600 nm, and the second overtones occur between 1000 and 1200 nm [11] whereas the aromatic CH absorptions of the UV stabilizer are shifted to shorter wavelengths [52]. As these bands were readily distinguishable, quantification of UV stabilizers containing aromatic CH absorptions was possible to a detection limit of 100 ppm. Percent slip is another typical additive measured by NIR in fabricated products such as polypropylene overcaps for bottles and or polypropylene container parts. NIR diffuse reflectance was used to determine Irganox 1010 and Irgafos 168 antioxidant content in a range of 0 to 4500 ppm in HDPE. Extraction of the antioxidants was performed either by microwave-assisted extraction (MAE) or ultrasonication and high performance liquid chromatography (HPLC) was used as a reference method. Multiple scattering correction (MSC) and second derivative spectral pretreatment were used to eliminate the size and shape physical differences of the samples. RMSEP for Irganox 1010 and Irgafos 168 was approximately 46 and 97 ppm, respectively [53]. NIR can be used to follow the influence of gamma-irradiation on the ageing of ethylene propylene diene monomer (EPDM) elastomers. PCA in the $5500\text{--}4500 \text{ cm}^{-1}$ or $7090\text{--}6980 \text{ cm}^{-1}$ range allowed discrimination and classification by increasing irradiation doses. This preliminary study opens up a cost-effective opportunity for the *in situ* and nondestructive monitoring of polymer degradations in nuclear power plants [54,55].

27.3.2 POLYVINYLCHLORIDE

Rigid PVC tiles were scanned to determine percent activator using an InfraAlyzer 500 Edapt-I [56]. This system transmits the light to the sample via low-OH bundled fibers. After striking the sample, the light is reflected into a gold-plated integrating sphere located on the end of the fiber probe. The activator in the PVC tiles varied from 0.08 to 0.4% and was measurable to $\pm 0.02\%$. Wavelengths used were 1268, 2066, 2150, and 2220 nm with an r^2 of 0.9813. The aromatic C—H stretching of the activator compound was shifted to shorter wavelengths [52] than the straight-chain C—H stretching of the PVC bands.

Plasticizer and processing aid content are measured in flexible PVC with this technique. Samples, typically coarse white powder, are poured into a closed cup with a low-moisture quartz window and a screw-close bottom, and scanned in reflectance mode. Plasticizer, dioctyl phthalate, ranges from

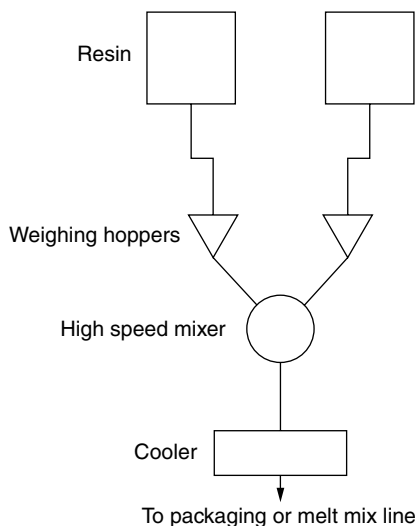


FIGURE 27.4 Schematic of a typical powder mix line. NIR on-line analyzers can be used to monitor addition of dry additives.

24.75 to 29.50% and is measurable with an r^2 of 0.994 and an SEE of $\pm 0.2\%$. The processing aid, expoxidized soybean oil, is determined with an r^2 of 0.992 and an SEE of $\pm 0.1\%$. The important bands were 1650, 1690, and 1700 nm. The wavelength at 1690 nm is an overtone of the aromatic C—H on the phthalate ring. The absorption band is shifted to a shorter wavelength than aliphatic C—H bands. The other wavelengths are most likely used to reference the absorbance of the PVC backbone. This application is well suited to an online application. The typical powder mix line is shown in Figure 27.4.

NIR reflectance spectroscopy was used to develop a nondestructive and rapid qualitative method for the analysis of plastic films used by the pharmaceutical industry for blistering. Three types of films were investigated: PVC, PVC coated with 40 g/m² of polyvinylidene dichloride (PVDC), and PVC coated with 5 g/m² of TE (Thermoelest) and 90 g/m² of PVDC. Analyses aimed at identifying one type of film and rejecting all false types (different thickness, color, or layer) were developed using different pretreatment options and PLS. True and false samples from four plastics manufacturers were included in the calibration sets to obtain robust methods that were independent of the supplier. Specificity was demonstrated by testing validation sets against the methods. The tests showed 0% of type I (false negative identification) and 1% of type II errors (false positive identification) for the PVC method, 13 and 3%, respectively, for the PVC-PVDC method and no error for the PVC-TE-PVDC method. Type II errors, mostly due to the slight sensitivity of the methods to film thickness, were easily corrected by simple thickness measurements [57].

27.3.3 POLYVINYL ALCOHOL

Polyvinyl alcohol (PVOH) has been analyzed for OH number in the solid form using NIR. The samples were scanned from 1100 to 2500 nm “as is” in a solid form using NIR. Hydroxyl number ranged from 86 to 89 and the SEE $\pm 0.2\%$. Key wavelengths included 1650, 1660, and 2062 nm. The 1650 and 1660 nm wavelengths are used to reference the —CH₂ backbone in the matrix and 2062 nm is an O—H combination band.

PVOH films have been analyzed for moisture and additives. The moisture in PVOH film gradually equilibrates with atmospheric conditions over time. The —OH in moisture hydrogen-bonds with the OH in the PVOH films.

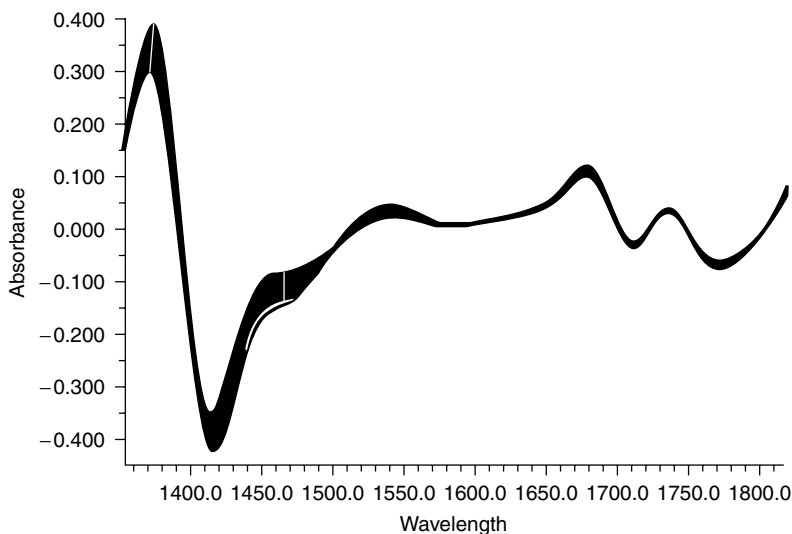


FIGURE 27.5 O—H band shift between 1400 and 1500 nm as PVOH is scanned over a period of 1 h. Calibrations for PVOH should be based on wavelengths that are more stable with time.

Hydrogen bonding shifts the frequency of the —O—H to longer wavelengths [11]. Bond shifting in the 1400 to 1700 nm and the 1850 to 2150 nm was intense in PVOH film. In addition, the first scan was dramatically different because both inter- and intra-hydrogen bonding were disrupted by the initial NIR beam energy incident to the sample. The film was scanned from 1100 to 2500 nm using a transfectance in a cell with a low-moisture quartz window and a screw-close back with a ceramic backing.

To demonstrate the wavelength shift in this region, an InfraAlyzer 500 was set up to scan continuously for 1 h (Figure 27.5). Note that the overtone region for —O—H changes with time and is unstable. This shift is due to a change in hydrogen bonding. There are areas, however, where the bonds are more stable with time. Although the overtone O—H region was selected as being statistically the best wavelength region, this area is not the best pick from a spectroscopic point of view. The region where the bands do not shift with time are more reliable for measurement of the moisture content of PVOH.

Pure PVOH films with molecular weights (5,000, 17,000, 72,000, and 1,25,000 gm/mol) were prepared by a casting technique to 0.21 mm thickness. The UV-vis-NIR absorption spectra gave the same band positions but the absorption intensity varied in a reverse trend with increasing molecular weight. The dependence of the absorption coefficient on the photon energy was determined and the band tails and energy gaps were calculated. It was found that the band tail increased while the optical gaps, Young's modulus, and the strength at the break for PVOH films decreased as the molecular weight increased [58].

PVOH was analyzed in synthetic pulp containing polypropylene and cellulose. The 2100-nm band could not be used to measure PVOH in the presence of cellulose, so a band in the 2350-nm region was used for PVOH analysis [59]. In the textile industry, PVOH is used for “sizing” textile materials as they pass through the plant and is eventually removed in the “desizing” step. NIR is becoming widely accepted as a quantitative tool for desize testing [60].

Vinyl acetate/vinylchloride copolymers were analyzed with diffuse reflectance and photoacoustic technique. The 2150-nm band, assigned as an ester carbonyl, was used [61]. Ethylene/vinyl acetate copolymer was quantified using an InfraAlyzer 500 and a rotating cup drawer. Vinyl acetate was

measured between 1.6 to 32.2% with an r^2 of 0.9999 and an SEE of ± 0.16 using wavelengths of 1500, 1548, and 1644 nm [13].

27.3.4 POLYVINYL ACETALS

Commercial polyvinyl butyrate is manufactured by reacting polyvinylacetate in alcohol to form a PVOH. An acid catalyst and butyraldehyde are then added to the PVOH to form polyvinyl butyrate, a free-flowing powder. The resulting hydroxyl concentration (expressed as a percent of PVOH) ranges from 17 to 22%.

A quality assurance method was developed for hydroxyl content. To maximize accuracy and precision, the white powder was milled in a Udy cyclone grinder, with a 1-mm screen and then packed into a standard closed cup. The calibration range of hydroxyl was 17 to 22%, with an SEE of $\pm 0.2\%$. An r^2 of 0.993 was obtained. A five-wavelength calibration using 1734, 1940, 2100, 2208, and 2270 nm was used.

The melt-extrusion transesterification of ethylene/vinyl alcohol (EVAL) copolymers was monitored by on-line NIR, measured at the exit of the extruder by use of a fiber-optic probe. Calibration models for predicting the vinyl acetate content in EVA copolymers were developed using PLS regression. The correlation coefficient and SEP were 0.96 and 0.85%, respectively [62].

Self-modeling curve resolution (SMCR) analysis was applied to on-line NIR spectra of the melt-extrusion transesterification of ethylene/vinylacetate copolymer. SMCR techniques include orthogonal projection analysis (OPA) and simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) [63].

27.3.5 POLYSTYRENE

Polystyrene, with a glass transition temperature (T_{gi}) of about 100°C, is clear crystalline polymer at temperatures below 100°C, yet an amorphous and easy to mold thermoplastic at temperatures greater than 100°C.

Crystalline polystyrene has a distinct NIR spectrum and is often used to calibrate NIR spectrophotometers for a linear wavelength attainment over a spectral range. Measurements are made at 1143.7, 1680.2, 2167.2, and 2306.8 nm to test the linearity of the spectrophotometer (Figure 27.6). Further tests have been developed to facilitate testing on a daily basis [64].

Quality assurance methods have been developed to detect the concentration of additives in stabilized polystyrene. One approach is to press the polymer pellets into a disk using a sample prep device located on most extruding machines. This disk is then placed on an optical flat and scanned from 1100 to 2500 nm using transreflectance. This reproducible sampling technique enables detection to 100 ppm of additive and greatly reduces prediction errors. Another method has been developed to measure ppm of phenolic antioxidant, ppm zinc stearate, and melt flow index in crystalline polystyrene and percent of rubber, ppm zinc stearate, and melt flow index in impact polystyrene. This method measures the polymer pellets whole, averaging multiple readings of the same sample in a rotating cup sample drawer [26].

Styrene-acrylic copolymers were analyzed via NIR for styrene content using the 2100 nm aromatic C—H combination band [65]. Spectral characteristics of these blends are so distinct that no special sample preparation technique is required. NIR has also been applied for the compositional analysis of styrene-acrylonitrile (SAN), styrene-MMA, styrene/butadiene [15], AS in polyvinylchloride [66], and styrene-acrylic copolymers.

Rubber-modified polystyrenes, such as ABS polymers, are two-phase systems in which the elastomer component is dispersed through the rigid phase. These rubber-modified styrene polymers can be analyzed directly in the latex (or aqueous emulsion) phase. Although aqueous content can approach 99%, the overtone and combination bands of the OH of water fall between 900 and 1000;

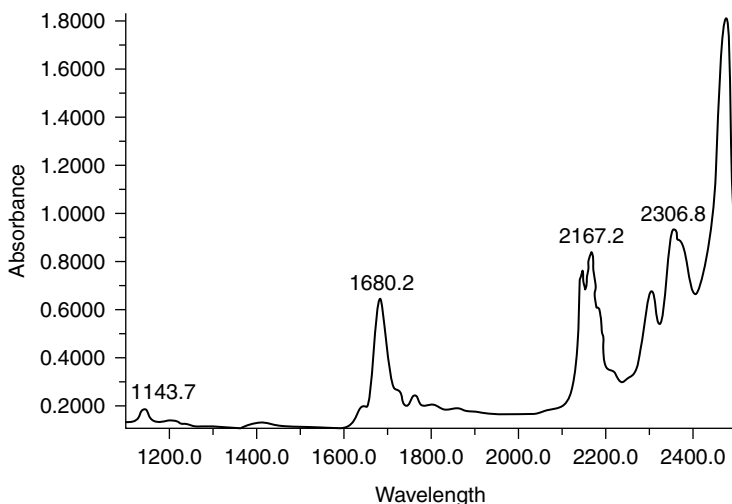


FIGURE 27.6 Four absorbance bands of crystalline polystyrene are used to test the linearity of wavelength attainment in NIR spectrophotometers.

1400 and 1500; and 1900 and 2000 nm. The O—H absorption bands are broad. However, the O—H bands are not present from 700 to 900; 1500 to 1900; and 2000 to 2500 nm. The overtone and combination bands of (C—H)_n-related bands are measurable in these regions.

Samples were poured directly onto a 0.3-mm cell specially designed for viscous materials. This pathlength improves the sensitivity in the C—H regions. Copolymer ratio and percent solids were determined directly on the sample, eliminating the need to dry down the sample.

Resin content in glass-filled polystyrene materials was analyzed without sample preparation. Analysis can be performed at line to control compounding and blending at the molding machine.

Tailored polymer resins are frequently required for a given application. Fontoura et al. used NIR spectroscopy for in-line and *in situ* monitoring and control of monomer conversion and polymer average molecular weight during styrene solution polymerization. Two process control strategies, one based on the optimal control theory and the other on model predictive control, were implemented both theoretically and experimentally [67].

Cross-linking was established by measuring percent water uptake, determined gravimetrically and by NIR spectroscopy in cross-linked sulfonated styrene-butadiene-styrene (SSBS) triblock copolymers. Cross-linking was through condensation of sulfonic acid groups with bis(4-aminophenyl)sulfone to form sulfonamide linkages using a coupling agent (carbonyldiimidazole) [68].

Generalized two-dimensional near-infrared (2D NIR) correlation spectroscopy was applied to the study of the conformational changes and specific interactions in blends of atactic polystyrene (PS) and poly(2,6-dimethyl-1,4-phenylene ether) (PPE). NIR diffuse reflectance spectra were measured for pure PS, PPE, and blends. The samples were divided into two sets for 2D correlation, a set with high PS content (70 to 100%) and a set with high PPE content (PS/PPE = 50 to 90%). The 2D synchronous correlation analysis separated the bands of PS from those of PPE. The 2D asynchronous correlation analysis produced cross-peaks that were indicative of the specific interaction or the conformational change in the blends. Two “blend bands” at 8443 cm⁻¹ (PPE) and 6887 cm⁻¹ (PS), were attributed to the molecular level changes induced by the formation of the blends. The 7700 to 6400 cm⁻¹ region was most suitable for 2D analysis of the specific interaction between PS and PPE, while the 5600 to 4900 cm⁻¹ region was best for 2D analysis of the conformational features common to both sets, and the 4750 to 4300 cm⁻¹ region was best for conformational features unique to the sets.

It was concluded that not only the phenyl rings of PS and PPE but also the CH₃ groups of PPE play important roles in the formation of the blends [69].

27.3.6 ACRYLIC AND METHACRYLIC POLYMERS

27.3.6.1 Ester-Based

Key analytical parameters in the manufacture of acrylic esters are residual monomer and residual OH content. Residual OH was analyzed in polymethacrylate. A corrosion-resistant cell with temperature equilibration at 15°C was used to reduce sampling error. Results were analogous for samples analyzed in transmission and transfectance. The range studied was 0.5 to 7.5% residual OH. An r^2 of 0.992 and an SEE of 0.3 were obtained.

Composition of methyl acrylate (MA) and methyl methacrylate (MMA) comonomer was analyzed in temperature-equilibrated a liquid cell (30°C). The MA concentration ranged from 10 to 18% and the MMA concentration ranged from 82 to 99%. The MA calibration used the following wavelengths: 1728, 2312, and 2400 nm; and the MMA calibration used 2192, 2280, and 2400 nm. Combination bands are affected by small molecular symmetry differences and thus are often used to separate materials with similar organic functional groups.

Other NIR applications include the measurement of the polymerization of MMA and the residual monomer in acrylic sheets [15]. ABS and acrylate latexes can be tested for composition and percent solids. Wavelengths used for percent butyl acrylate were strictly CH₄ combination bands whereas those used for percent solids (1680, 1710, 2270) represented both CH_n overtone and combination bands [11]. The butyl acrylate ranged from 0.2 to 22%, with a SEE of ± 0.1 and an r^2 of 0.98. Solids ranged from 16 to 18% with a SEE of ± 0.1 and an r^2 of 0.99. Miller and Yin [70] studied polyoctadecylmethacrylate absorbed on alumina by NIR.

NIR and Raman were used to analyze polymer brushes with single-walled carbon nanotubes (SWNT) as backbones. Carboxylic acid groups on SWNT were formed by nitric acid oxidation. Atom transfer radical polymerization (ATRP) initiators were covalently attached to the SWNT by esterification of 2-hydroxyethyl 2'-bromopropionate with the carboxylic acid groups. *n*-butyl methacrylate (nBMA) was grafted from the ends and sidewalls of SWNT via ATRP. Methyl 2-bromopropionate (MBP) was added as free initiator during the brush preparation to control growth of the brushes and to monitor the polymerization kinetics. NIR and Raman spectra indicated that the side walls of the SWNT were lightly functionalized by the nitric acid treatment and that the degree of functionalization of the SWNT did not change significantly during the formation of initiator or during the polymerization [71].

Fiber-optic NIR spectroscopy was successfully used to monitor the conversion of monomer during the solution polymerization of MMA carried out in a lab-scale reactor. NIR spectra were recorded during 18 batch and semi-continuous reactions using an *in situ* transmission probe. PLS was used to relate the NIR spectral data to the conversion measured off line by gravimetry. The calibration had a correlation coefficient of 99.45% and a standard error of calibration (SEC) equal to 1.95% and was validated for various operating conditions. The conversion was measured with a standard error below 2.6%. The on-line method was demonstrated to be accurate, robust, and suitable for versatile use in the polymerization plant [72]. NIR was also used to determine the weight-average molecular weight using an *in situ* transmission probe with GPC used as reference method. The measurements were validated for various operating conditions including different solvent, initiator, MMA, and chain-transfer agent concentrations, and for both batch and semi-continuous modes. The on-line evaluation of MW was obtained with an average relative SEP of 5.7%; such on-line NIR measurement was thus demonstrated to be robust and accurate, even for varied experimental conditions [73].

ESR and NIR were used together to obtain reliable values for the propagation rate coefficients as a function of conversion for polymerizations of MMA and a mixture of MMA and ethylene glycol

dimethacrylate (EGDMA) with a 1:1 concentration of double bonds. ESR spectroscopy was used to measure the radical concentration while FT-NIR fiber-optic spectroscopy was used to measure instantaneously the double-bond concentration within the temperature-controlled cavity of the ESR instrument during polymerization. The advantage of this approach for measuring rate coefficients is that it is equally applicable to branching and linear polymerizations [74].

NIR was used for the simultaneous in-line and *in situ* monitoring of monomer (MMA and butyl acrylate, BuA) and polymer concentrations in the reaction medium during seeded semibatch emulsion copolymerizations. Calibration samples were selected from several reactions, and PLS was used to build three independent calibration equations in the range of 1100 to 1900 nm [75].

The CH stretch NIR spectral features were assigned by a comparison of the number and type of different CH oscillators in ethylene glycol diacrylate (EGDA) and EGDMA monomers polymers and a comparison to the spectra of acrylic acid and ethylene glycol disobutyrate. Peaks corresponding to transitions for the vinyl, methylenic, and methyl types of the CH stretch were identified. From the integrated areas of the vinyl and methyl/methylenic peaks for the first, second, third, and fourth vibrational overtone transitions of EGDMA, the degree of polymerization was calculated and compared well with the degree of polymerization determined from MIR spectroscopy [76].

NIR and dynamic mechanical analysis (DMA) to determine glass transition temperature (T_g) were used to investigate whether a material with a high-glass transition temperature can be made from a polymerization performed at a significantly lower temperature (e.g., ambient temperature). To understand the factors that play a role in determining the relationship between T_g and cure temperature (T_{cure}), three model systems with different degrees of heterogeneities glycol dimethacrylate (TEGDMA), diethylene glycol dimethacrylate (DEGDMA) (more heterogeneous), and styrene-divinyl-benzene copolymer (St-co-DVB) (more homogeneous). These systems were photopolymerized in a temperature cell while simultaneously monitoring the sample temperature with the T_{cure} varying from 25 to 95°C. NIR was used to monitor the polymerization rate and final double-bond conversion. The T_g of the exact sample cured within the NIR was subsequently measured using DMA. Since this work utilized living radical photopolymerizations, the traditional issues associated with characterizing chain polymerizations (specifically, trapped radicals that persist after cure) were eliminated. It was found that the differences between T_g and T_{cure} were significantly greater for the more heterogeneous multi DEGDMA system than for the less heterogeneous St-co-DVB system with the $T_g - T_{\text{cure}}$ being as large as 100°C. Therefore, heterogeneous networks with broad distributions of relaxation times like DEGDMA exhibit unique cure behavior and facilitate obtaining a higher T_g as a function of T_{cure} than is possible in a comparable, more homogeneous network [108].

Polymer particle size could be estimated from NIR spectra collected in situ and online in a automated reactor. A good linearity between spectra and average polymer particle diameter was found. The multivariate linear model was also tested with an independent data set not used in the model fitting. Despite varying monomer and polymer concentrations and temperature changes along the process, the average particle sizes were well predicted and could be monitored by NIR spectroscopy during emulsion polymerization reactions [109].

27.3.6.2 Amide-Based

NIR has been used to determine total nitrogen in acrylamide cellulose esters [85]. Nitrogen content can be correlated to “crosslinkability” of these polymers. Manual results were obtained using the Kjeldahl procedure. Nitrogen content ranged from 0.3 to 1.3% and SEE of the NIR method was ± 0.02 with an r^2 of 0.998. Samples were coarse powders and were scanned as is in a standard closed cell from 1100 to 2500 nm.

Moving window two-dimensional (2D) correlation spectroscopy, which partitions a data set into series of relatively small submatrices (windows) and calculates their covariance maps in succession,

was tested for phase-transition temperatures of poly-(N-isopropylacrylamide) in an aqueous solution. Covariances of the temperature-dependent NIR and infrared spectra, respectively, were investigated to find the spectral differences and to classify randomly ordered spectra by a sample-sample covariance map. The criterion of mean of standard deviation of covariance matrices was used as an indicator of the crucial information present in these matrices. The baseline variation was not removed prior to the calculations but was found to be of use for the determination of the phase-transition temperatures [110].

The principles of RAFT polymerization were applied to the polymerization of N-isopropylacrylamide (NIPAAm), carried out in the presence of the dithiocarbamates benzyl 1-pyrrolicarbodithioate or cumyl 1-pyrrolicarbodithioate as chain transfer agents in 1,4-dioxane at 60°C. A kinetic investigation using *in situ* FT-NIR spectroscopy showed very long induction periods, which depended on the nature and concentration of the chain transfer agent. The resulting polymers had polydispersity indices $M_w/M_n < 1.3$ and were analyzed by MALDI-TOF mass spectrometry, GPC, NMR, and UV spectroscopy [111].

27.3.7 POLYAMIDES

A majority of polyamides go into manufacture of textile fibers. Ghosh and Rodgers used a filter analyzer to determine heat set temperature of nylon yarn [23]. Their work was based on the ratio of crystalline to amorphous nylon. Combination bands were vital in resolving the spectral characteristics of crystalline to amorphous nylon ratios. Percent moisture in caprolactam was studied for eventual placement online. The laboratory study measured moisture in the range of 0.05 to 9% with a standard error of prediction (SEP) of 0.4 and an r^2 of 0.995. Samples were scanned from 800 to 2500 nm at room temperature with satisfactory results. Improved accuracy and precision are attainable by use of a temperature correction algorithm.

Additives to thermoplastic nylon were analyzed on a fixed filter analyzer. The samples were presented as cubes and films with both presentation methods giving similar results. The additive range was 0.25 to 1.25%, with an SEE of ± 0.04 and an r^2 of 0.99.

NIR transmission spectroscopy and thermogravimetric analysis (TGA) were used to predict diffusion of water into polyamide 6,6 samples immersed in water at varied temperatures and time. The diffusion was time dependent. The absorption curves were fitted by the use of a time-dependent surface concentration and a water-concentration-dependent diffusivity. The water diffusivity decreased non-linearly and the activation energy of diffusion increased from 24 to 58 kJ mol⁻¹, with decreasing temperature. The accuracy and feasibility of this method was similar to conventional TGA methods, however FT-NIR provided the possibility of monitoring the moisture concentration on-line and non-destructively [112].

The diffusion of deuterium-labeled (OD) low-molecular-weight alcohols and D₂O in Nylon 11 was investigated by monitoring the NH/ND exchange in the polymer by FT-NIR spectroscopy. The results demonstrated that the diffusion process of the different penetrants was strongly controlled by their molecular structure and geometry. To characterize this phenomenon quantitatively, the diffusion coefficients were determined. The rate of the diffusion was found to decrease with increasing size of alcohol. Furthermore, it could be shown that with this technique the less ordered regions of the polymer could be separated spectroscopically from the crystalline domains, which were not accessible for the isotopically labeled diffusants [113].

De Araujo and Y. Kawano used curve fitting calculations and local mode theory to assign overtone and combination frequencies in the NIR spectra of polyamide 6 (PA 6), poly(vinyl chloride) (PVC), and polychlorotrifluoroethylene (PCTFE). Anharmonicity correction and mechanical frequency were determined from a Birge-Sponer plot. Anharmonicity corrections of 55, 61, and 20 cm⁻¹ were obtained for CH₂ NH, and CO stretch modes of PA 6, respectively, and of 60 and 66 cm⁻¹ for CH₂ and CH stretch modes of PVC, respectively. The local mode model seemed to be adequate to interpret the origin of the bands of PA 6 and PVC. Anharmonicity corrections of 33, 19, and 16 cm⁻¹ were

obtained, respectively, for CF, asymmetrical CF₂, and symmetrical CF, stretch of PCTFE functional groups [114].

The kinetics and mechanisms of thermally initiated (using 2,2'-azobisisobutyronitrile (AIBN) as initiator) radical homopolymerizations of a series of maleimides, including N-phenylmaleimide (PHMI) [1-phenyl-1H-pyrrole-2,5-dione]; N-n-hexylmaleimide (nHMI) [1-(n-hexyl)-1H-pyrrole-2,5-dione]; and N-cyclohexylmaleimide (CHMI) [1-cyclohexyl-1H-pyrrole-2,5-dione] were investigated in THF solution by an on-line FT-NIR technique. It was found that the order of the activation energies for the three N-sub-MIs is: E-a PHMI < E-a (PHMI) < E-a (CHMI). The overall polymerization rate parameter *k* and the pre-exponential factor *A* were calculated. The kinetic order with respect to the N-sub-MIs was in the range of $0.71 < m < 0.75$ for the initiator and $n = 1.0$ for the monomer. Radical transfer to solvent was found to be the key factor in determining the apparent order with respect to the initiator. All of the homopolymers had a relatively low molecular weight. The end groups of the polymer chains were characterized by MALDI-TOF, GPC and NMR methods and the results indicated that the polymerization was initiated by THF radicals, and that the termination reaction was mainly controlled by chain transfer to solvent through an hydrogen abstraction mechanism [115].

27.3.8 CELLULOSE ESTER

The degree of substitution (DS) or acetyl value of cellulose acetate is a measure of the degree of esterification of cellulose. The manual test involves saponification followed by a back titration. NIR has been used to analyze acetyl value by measuring the hydroxyl number directly and then the acetyl value from the OH number results. Mitchell et al. [116] studied acetyl content in cellulose acetate by NIR as early as 1957. De Wit et al. published their work on routine testing of cellulose esters [117]. Plasticizer content has been analyzed in a cellulose acetate by a fixed filter analyzer. The plasticizer ranged from 9.8 to 23.1% with an SEE of 0.4% and an *r*² of 0.994. Wavelengths used were 1680, 1722, 2336 nm (related to -C-Hn stretching) and 1982 (related to the second overtone of C = O).

An NIR method was developed for the simultaneous determination of percent polymer and moisture in a wood pulp substrate. Samples are placed directly on the optical window of the filter analyzer. Results are reported as follows:

	Polymer	Moisture
Range, %	0 – 22	5 – 6
No. of Samples	20	20
<i>r</i> ²	0.999	0.996
SEE	0.4	0.5

27.3.9 SUBSTITUTION ON STARCH

NIR was used to analyze hydroxyethyl substitution on starch. The range was 1.0 to 2.5% with an SEE of 0.15%. The sample form was a coarse white powder, and was well suited to the standard closed cup. Table 27.1 shows the comparison of manual and NIR values for percent hydroxyethyl substitution on starch.

NIR can be used for quantitative analysis of the degree of substitution of octenylsuccinic anhydride, acetyl, and propylene oxide as well as protein contamination.

TABLE 27.1
Comparison of Manual and
NIRA Values for Percent
Hydroxyl Ethyl Substitution
on Starch

Manual	NIRA	Difference
1.06	1.09	−0.03
1.72	1.69	0.02
2.17	2.15	0.02

27.3.10 ALCOHOLS

Analysis of the —OH functional group band is one of the major applications of this technology. Manual —OH determination usually involves a phthalation and back titration, which is time-consuming and requires the use of noxious chemicals. Hydroxyl number is defined as the mg of KOH equivalent to the hydroxyl content of 1 g of alcohol. Early work by Hilton [118] demonstrated the utility of the NIR region for measuring hydroxyl group absorption bands. LeFevre et al. [119] and Miller and Willis [15] used the —OH bands near 1400 and 2100 nm to determine the molecular weight of polyethylene glycols. Filter-based analyzers have been used for this application due to their inherent ruggedness and minimal maintenance requirements. No a priori knowledge on the part of the process operator is required, yet the analyzer provides accurate and precise results. C. Jones and Brown [120] monitored hydroxyl number in polyols using a single-filter single reference spectra photometer. Byron [121] and Turley and Pietrantonio [122] used filter analyzers for simultaneous hydroxyl number and moisture. Ethylene oxide was also determined. Online systems are also in use for this application in combination with fiber-optic sampling systems. Grob et al. describe remote online monitoring of polyol production using NIR [123]. Tan et al. developed an improved standard-free method, finite impulse response, for NIR hydroxyl number calibration transfer for terpolymers of poly(vinyl chloride), vinyl acetate, and hydroxypropyl acrylate with range of 1.38 to 1.83 OH number [124].

27.3.10.1 Fatty Alcohols (Higher Aliphatic Alcohols)

In the fatty alcohol process, NIR is used for raw-material quality testing to check residual free fatty acids and moisture content. In the hydrogenolysis reaction, conversion of the ester or acid to alcohol can be monitored. NIR is also used as a quality assurance tool for the final product. A fatty alcohol calibration was developed on a fixed filter instrument which included the key wavelength 2040 nm. Samples were injected into a fixed-pathlength liquid transfectance cell with temperature set at 45°C. This sampling design provides a high level of accuracy and precision. The OH number range was 150 to 430. The SEP was ± 0.4 , with an r^2 of 0.99999, and an F value of 99,255. When an analytical technique is used online, absolute accuracy can sometimes be “traded-off” for ease of sampling because of the frequency of measurement. Typical lab techniques measure only a few results to represent an entire batch where an online measurement can generate results every few seconds. Plotting these results in statistical control charts shown in Figure 27.7 [125] gives a better profile of the product target versus upper and lower limit. In this case, the temperature equilibration step was eliminated when the method was transferred online for process control of OH number in fatty alcohol manufacture. As a result there was a small increase in the error of the method, but the calibration performed well enough with respect to the required accuracy and precision to be

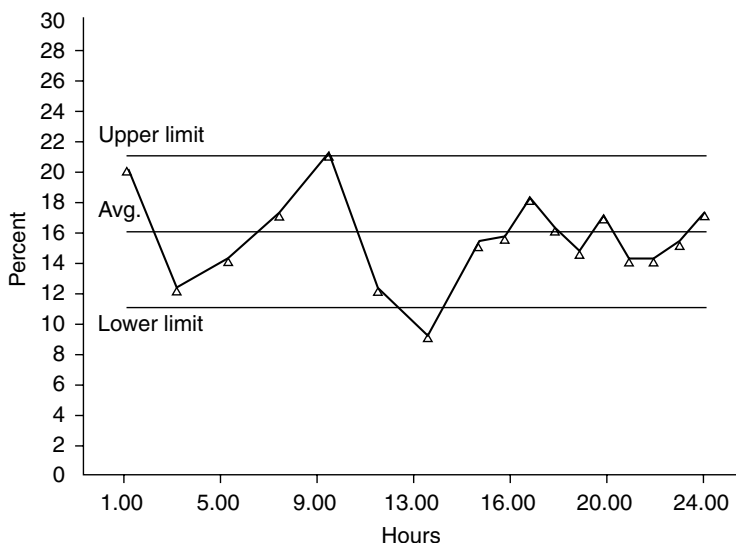


FIGURE 27.7 When using an NIR method for process control, statistical control charting can be used to set upper and lower control limits. This method clearly shows the difference between one sample with an unusual result vs. a series of samples that are trending toward the upper or lower control limit.

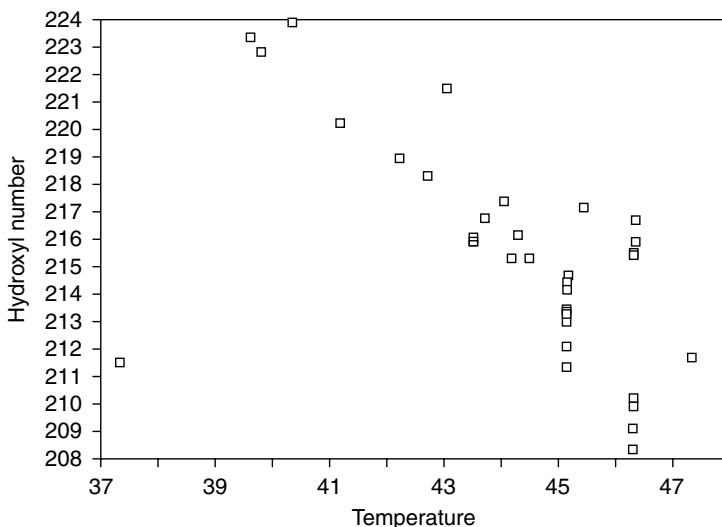


FIGURE 27.8 A linear trend is observed between temperature and predicted hydroxyl number. Temperature can be included in the calibration as an indicator variable or via a temperature correction algorithm.

implemented. In this case, the benefits of frequent testing and ease of sampling were a consideration in the calibration development step.

The temperature sensitivity of fatty alcohols has been investigated. Samples were injected into a variable transfectance temperature cell. The temperature on the cell was varied from 37 to 45°C and sample measurements were taken at 1°C intervals. A linear trend is observed between temperature and predicted hydroxyl number (Figure 27.8). Over a range of approximately 2°C, it appears that the error due to temperature effects is tolerable.

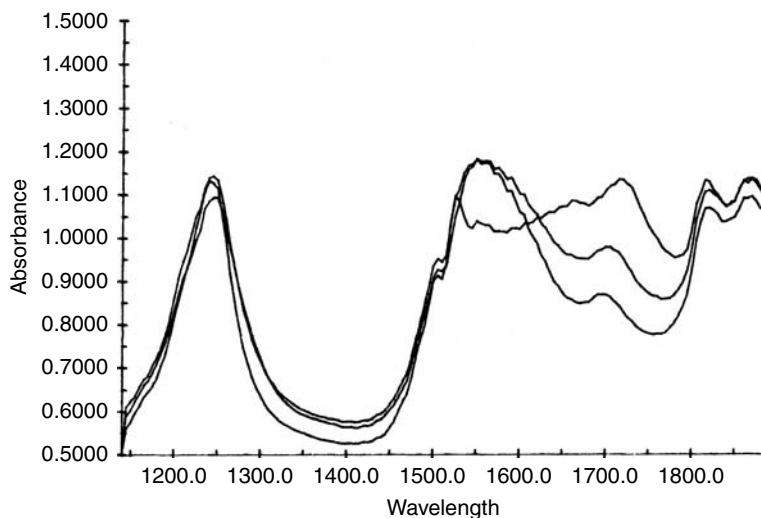


FIGURE 27.9 Hydroxyl number can be determined online using AOTF technology, which enables ultrafast scanning with no moving parts.

Temperature can be used as an indicator variable [94] or a temperature correction algorithm can be used to compensate for temperature variations greater than 2°C. The same samples were analyzed in transmission, and the results were analogous.

27.3.10.2 Synthetic Alcohols

NIR is used in both process control and quality assurance of synthetic alcohol manufacture. A calibration was developed on a precision scanning analyzer using a temperature-controlled liquid cell at 45°C. Samples ranged from 130–290 OH. The r^2 was 0.9998, with an SEE of ± 0.8 OH and an F value of 14,000.

An InfraPrime AOTS system was used to analyze synthetic alcohols as shown in Figure 27.9. Ultrasonic waves are used to tune the light energy, resulting in a spectrophotometer with no moving parts and flexible, ultrafast scanning. The spectra in the figure were obtained by scanning through a disposable glass vial. However, fiber-optics sampling systems are most commonly used to collect the spectra.

The unique capability of NIR/IR spectroscopy to characterize the hydrogen bonding of the OH groups in ethylene-vinyl alcohol copolymer (EVOH) was demonstrated. FT-NIR/IR spectroscopic measurements were made of EVOHs with the ethylene content of 33, 49, 75, and 89 mole %. A strong, broad absorption at ~ 4780 cm^{-1} , which is the first combination of the OH stretching and deformation vibrations, was assigned to the aggregated OH groups. The band shows a shoulder at 4970 cm^{-1} at the ethylene content of mole 49%, which became a distinct peak at 89% content. This was assigned to the free OH group isolated in the surroundings of the ethylene component. Existence of the free OH group was confirmed by the appearance of sharp absorptions of the OH stretching fundamental at 3027 cm^{-1} and its first overtone at 7094 cm^{-1} . The relative population of the free OH was estimated as 0, 0.3, 1.5, and 3.9 mole % for the ethylene content of 33, 49, 75, and 89 mole %, respectively [126].

27.3.10.3 Polyhydric Alcohols

NIR analysis of sorbitol was attempted in aqueous solutions. Problems arose because mannitol, an isomer of sorbitol, was also present in solution.

The two were indistinguishable in aqueous solutions. Mannitol and sorbitol, however, are crystalline as a final product and were readily distinguished in the solid form with NIR [127].

27.3.10.4 Esterified Alcohols

A method to determine hydroxyl number of sorbitan esters has been developed using a scanning NIR analyzer. Hydroxyl number ranged from 7.4 to 329, with an SEE of ± 11 and an r^2 of 0.0071. The SEE was improved when the teaching set was divided into subclasses. The ranges of OH number on the two subclasses were 0 to 50 and 51 to 329.

27.3.10.5 Ethoxylates

NIR has been used to measure the percent of water in ethylene glycol, one of the raw materials used in the production of polyethoxylates. Samples were scanned from 1100 to 2500 nm in a thermostated liquid cell (45°C). An SEE of $\pm 0.15\%$ was obtained over the range 0 to 10% H₂O. The moisture combination band, 1940 nm, was used as well as a (CH)_n combination band in a two-wavelength calibration.

Molecular weight of polyethylene glycol can be calculated by dividing the measured hydroxyl number by 34 (Molecular weight = 34/OH). A typical calibration ranges from 31-137 OH number with an SEE of ± 0.15 and an r^2 of 0.998. Samples were run on a filter instrument in a 65°C thermostated transmittance cell. Table 27.2 shows lab versus Infra Alyzer-predicted results for polyethylene glycol molecular weight determination and OH determination.

A calibration for hydroxyl number and moisture was developed using 43 samples, all PO and EO-PO (heteric and block) diols, triols, and blends [122].

Samples were analyzed on a filter system using 65°C thermostated liquid cell. Results are summarized in Tables 27.3 and 27.4. SEE can be improved if calibrations are generated for specific ranges.

Ethylene oxide (EO) was analyzed as a single variable and in combination with glycerin (EO+GLY) [122]. Results are summarized in Table 27.5.

FT-NIR fiberoptic spectroscopy was used to monitor the anionic polymerization of ethylene oxide (EO), and the influence of different initiators and reaction conditions on the kinetics of polymerization. NIR online monitoring revealed an unexpected induction period present in EO homopolymerizations as well as in the synthesis of PEO containing block copolymers with [Li/(BuP4)-P-t](+) counterions. The resulting polymers were characterized by size exclusion chromatography (SEC). A low-molecular-weight polystyrene-block-poly(ethylene oxide) (PS-b-PEO) diblock copolymer was synthesized to provide more insight into the observed induction period by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-ToF MS) on samples taken during EO polymerization. The induction period was attributed to different factors involved in the formation of active centers, for example, the break up of lithium alkoxide aggregates by the phosphazene base (BuP4)-P-t, and chain length effects. It also depended on reaction temperature, concentration of the phosphazene base, as well as the structure of the initiator [128].

27.3.10.6 Alkoxy Alcohols

Alkoxylates of various types have been analyzed by NIR, including alcohol and nonylphenol ethoxylates, polyethylene glycol, polypropylene glycol, and copolymers of various base materials [122, 129, 130].

A polypropylene oxide of triethanolamine was studied. This product has a distribution of zero, mono, di, etc. adduct. The OH number obtained by NIR is an average of the various types of -OH groups with slightly different absorptivities at one wavelength. In order to distinguish specific types of OH groups contributing to the overall OH number, reference data must be obtained by NMR

TABLE 27.2

Lab vs. InfraAlyzer Predicted Results for Polyethylene Glycol Molecular Weight Determination and OH Determination

Polyethylene glycol MW determination ^a			Polyethylene glycol —OH determination ^b		
Lab	InfraAlyzer	Difference	Lab	InfraAlyzer	Difference
316	307	−9	10.77	11.06	0.29
316	305	−11	10.77	11.13	0.36
308	308	0	11.04	11.05	0.01
308	308	0	11.04	11.04	0.00
387	410	23	8.78	8.30	−0.48
387	409	22	8.78	8.32	−0.46
201	202	1	16.94	16.86	−0.08
201	203	2	16.94	16.74	−0.20
204	198	−6	16.64	17.20	0.56
204	198	−6	16.64	17.20	0.56
393	407	14	8.65	8.36	0.29
393	405	12	8.65	8.40	−0.25
195	192	−3	17.47	17.67	0.20
195	191	−4	17.47	17.81	0.34
195	191	−4	17.47	17.82	0.35
400	408	8	8.50	8.34	−0.16
400	405	5	8.50	8.39	−0.11
196	198	2	17.39	17.18	−0.21
196	197	1	17.39	17.26	−0.13
305	303	−2	11.13	11.23	0.10
305	304	−1	11.13	11.19	0.06
313	316	3	10.85	10.77	−0.08
313	315	2	10.85	10.78	−0.07
406	414	8	8.37	8.21	−0.16
406	414	8	8.37	8.22	−0.15

^a SEP = 8.520.

^b SEP = 0.283.

TABLE 27.3

Calibration Results with Wide Ranges of Hydroxyl Number

	Hydroxyl number	Moisture
Range	27–265 mg KOH/g	0.01–0.35%
r^2	0.9997	0.9761
SEE/ n^*	1.24/41	0.02
F	30,119	292

* n = number of samples in the calibration set.

Source: Data from Le and Hill, *Polym. Int.*, 2003. **52**: 1694–1700.

TABLE 27.4
Calibration Results Obtained
with a Tighter Range of Hydroxyl
Number

	Hydroxyl number
Range	25.6–111.1
r^2	0.9996
SEE	0.47
F	26,000

Source: Data from Le and Hill, *Polym. Int.*, 2003. **52**: 1694–1700.

TABLE 27.5
Comparison of Ethylene Oxide Analyzed as a Single Variable and in
Combination with Glycerin

	% EO + GLY		% EO	
	Low GLY	High GLY	Low GLY	High GLY
Range, %	0–73.1	0–73.1	0–71.6	0–71.6
r^2	0.9996	0.9993	0.9996	0.9996
SD	0.56	0.8	0.59	0.67
F	75,400	29,600	40,600	39,500
Wavelengths	1680	1680	1680	1680
	1790	1790	1722	1722
		2250	2100	1818
			2100	

Source: Data from Le and Hill, *Polym. Int.*, 2003. **52**: 1694–1700.

or HPLC and then correlated to the spectra using the data analysis programs available with NIR instruments [131].

27.3.10.7 Sulfonated Alcohols

Sulfonated alcohols are used for anionic detergents. Ammonium lauryl sulfate and sodium lauryl sulfate were analyzed for active, anionic detergent; solids; moisture; benzoic acid viscosity; and pH [132–134].

Linen/phenolic	Resin content	Flow content
Range	57–61	24–35
Number of samples	11	11
r^2	0.928	0.909
SEE	0.4	1.7
Wavelengths	1114	2083
	1422	2190
		2250

27.3.10.8 Method for Analyzing a Wide Variety of Alcohols

When many different types of products are being analyzed, a practical technique often utilized is to combine qualitative analysis with quantitative analysis. First, the unknown are classified by qualitative analysis, in other words, discriminant analysis [35,77]. Cascade software [78] automatically classifies a sample using discriminant analysis and steps to the proper calibration for quantitative determination.

For example, different calibration equations may be developed for sucrose-based polyols than for amine-based polyols. The instrument can automatically predict the unknown as belonging to the “amine-based polyol” class of compounds, then automatically predict the actual hydroxyl and moisture value. The analysis protocol is determined by both the accuracy requirements of the analyst and the diversity of materials to be routinely analyzed. This approach is very useful when analysts run a variety of product types.

27.3.11 POLYURETHANES

An NIR method was developed to measure moisture in the drying step of a polyurethane prepolymer process. Decreasing moisture content over time was apparent in the spectra [26].

A hydroxy-terminated polyurethane, prepared from ethylene glycol (EG) (0.5 mol) and 2,4-toluene diisocyanate (TDI) (0.3 mol), exhibits an N—H band at 1470 and 2050 nm. The 2050 nm band undergoes a large increase upon heat curing [16].

An isocyanate-terminated polyurethane, derived from TDI and polypropylene glycol with MW = 600, shows N—H bands at 1470 and 2040 nm. The 2040 nm band increases on heat curing [16]. NIR can be used to monitor the ratio of polyol or polyester being reacted with isocyanate. A substantial reduction in scrap rate occurs when precise ratios of polyol and isocyanate are known by the molding operator.

Miller et al. [79] analyzed rigid polyurethane foams by NIR spectroscopy. Further studies by Miller et al. [80–82] included the use of factor analysis to analyze polyester urethane urea block copolymers, and the determination of physical properties of reaction-injected-molded polyurethane by NIR-FT-Raman spectroscopy.

NIR transmittance spectroscopy was used to monitor polyurethane reaction extent *in situ* during a reaction injection molding (RIM) process. Good agreement was observed between multivariate (PCA) analysis, univariate (Beer–Lambert) analysis, and kinetic predictions. It was demonstrated that PCA methods with carefully chosen spectral ranges could provide estimation of the reaction extent without time-consuming calibration [83].

27.3.12 POLYESTERS

27.3.12.1 Alkyds

Kienle and Henry [84] showed that when phthalic anhydride and glycerol are heated together, the acid number drops very quickly as the dibasic acid reacts with primary hydroxyls. Phthalic anhydride then reacts with secondary hydroxyls to form a network structure.

Process control of alkyd manufacture depends on rapid measurement of acid number and viscosity. Acid number is expressed as the milligrams of KOH required to neutralize free fatty acid in 1 g of alcohol. Acid number and viscosity are plotted to determine the proper end point. When the resin has achieved the desired end point, it is pumped to a thinning tank.

NIR has been used to simultaneously determine acid number and hydroxyl number of alkyds. The simultaneous analysis of acid and hydroxyl number allows process operators to determine if the stoichiometry of the reaction is correct. Alkyds were analyzed at line by grabbing a sample, pouring it directly into a 0.5-mm viscous cell, and then placing it in a NIR filter analyzer. The viscous cell consists of a stainless steel base with a stage of a specified pathlength. Several drops of the sample are

placed on the stage, and a quartz cover glass is placed on the top. Excess sample runs into overflow troughs. The following summarizes the results:

Alkyd resin	Range	SEE	r^2	# Samples
OH #	40–180	4.1	0.99	34
Acid #	3–45	3.6	0.95	34

Advances have been made in fiber-optic sampling techniques that overcome problems with sampling high temperature and high viscosity materials. As a result, *in situ* sampling with fiber-optic probe systems is being implemented for these types of applications.

Crandall et al. [16] reacted phthalic anhydride and EG to show a decrease in the 1910 nm band due to C=O. A similar decrease in the 1910 nm band due to C=O was observed during the reaction of phthalic anhydride and glycerol. Viscosity, an important process parameter, has been quantified in alkyd resins using NIR. Qualitative NIR can also be used before the reaction begins to determine if the kettle has been “charged” properly. Discriminant analysis can be used to determine if all of the reactants are present and if they are present in the correct proportions.

27.3.12.2 Unsaturated Polyesters

Unsaturated polyesters are analyzed by NIR for hydroxyl and acid number [85]. Samples were poured into a 0.5-mm viscous cell (previously described), cooled to room temperature, and placed in an NIR filter analyzer. Results of two different unsaturated polyester calibrations are summarized in the following text:

Polyester	Range	SEE	r^2	# Samples
No. 1				
OH #	14–45	0.5	0.999	50
Acid #	1–38	0.5	0.999	50
No. 2				
OH #	15–70	0.8	0.998	20
Acid #	1–30	0.6	0.997	25

NIR is also being used to measure the binder and loss on ignition (LOI) of unsaturated polyester-fiberglass composites.

27.3.12.3 Saturated Aliphatic Polyesters

Hydroxyl number, acid number, and moisture in saturated aliphatic polyesters can be simultaneously determined using NIR.

27.3.12.4 Aromatic Polyesters

Production of high molecular weight PET for bottles and tire cord applications can be monitored using NIR. Carboxylic acid end groups can be generated as a by-product and, because of their acidity, can catalyze further degradation of the polymer. Thus, acid and hydroxyl number content are critical parameters. NIR has been used to measure acid and hydroxyl number at various stages of the process. White-powdered samples were ground in a Krupp mill and poured into a standard cup to improve

the accuracy and precision. Averaging and math treatments can be used to eliminate the need for grinding. However, one can expect slightly higher SEEs than those obtained in this study. Results for the simultaneous analysis of acid and hydroxyl number in PET is summarized in the following table.

Polyester	Range	SEE	r^2	# Samples
Acid #	2.5–75	3.0	0.99	29
OH #	52–118	3.2	0.98	16

Moisture analysis, also measurable by NIR, is another important parameter because the polyesters are typically dried to less than 200 ppm to prevent hydrolysis.

NIR has been used to measure percent degradation and viscosity of polyester terephthalate films. A method has been developed [86] to eliminate the effect of interference fringes and orientation due to stretching of the film. A special integrating sphere with PbS detectors at 0 and 90 has been used in combination with a sample cup using frosted quartz windows.

Miller and Eichinger [87] determined the crystallinity and morphology of fibrous and bulk PET by NIR reflectance spectroscopy. NIR applications for polyester fiber, such as percent finish, are discussed by Ghosh in the chapter on textile analysis in this textbook.

The initial oligomerization of bis(hydroxyethyl terephthalate), BHET, was monitored *in situ* by NIR spectroscopy with a transmission probe. The amount of hydroxyl end groups, a function of degree of the oligomerization, and the amount of EG in the reaction solution were successfully predicted by the NIR spectra combined with PLS regression using measured values from ¹H-NMR. Although the amount of EG was small, it was possible to extract spectral information about EG by use of loading plots [88]. This system was also monitored *in situ* at different temperatures simultaneously by attenuated total reflection (ATR)/infrared (IR) spectroscopy. The spectra were analyzed by generalized 2D correlation spectroscopy. Both 2D-IR and 2D-NIR correlation spectra clearly identified bands arising from EG, which was expelled out of the system during the reaction. It was found that NIRS may be very useful for *in situ* monitoring of the amount of water [89].

The influence of the surface roughness on the optical transmittance and on the interference spectral fringes produced by reflection of parallel layers of multilayer polymer films was analyzed using UV-Vis, NIR and IR spectra. Samples studied were multilayer materials for graphic arts, obtained by spreading adherent emulsion layers over both sides of a substrate of PET film with different surface roughness. Strong correlations between spectral parameters of the optical transmittance and surface roughness were found, showing that these spectroscopic techniques were good, complementary ways to study the surface characteristics of these engineering materials [90].

27.3.13 EPOXY RESINS

Epoxy resins contain two or more epoxide rings per molecule. The epoxy groups react with curing agents in a ring-opening reaction to yield high-performance thermosetting plastics. The reactivity of epoxy resins can be expressed as equivalent weight, which is the molecular weight per reactive group.

Epoxy resins are used for protective coating and in structural applications. When properly formulated and cured, they have excellent resistance to moisture and organic solvents, stable electrical characteristics under various environmental conditions, and an excellent adhesion to most materials. When reacted with curing agents, epoxies have very low shrinkage because no by-products are given off in their curing reaction. Determination of the state of cure advancement provides information on the nature and characteristics of polymerized materials. It is generally determined by the nature of

the hardness, film thickness interdependence, type of substrate, time, temperature, and the chemical formulation of the system. The traditional thumb test, scratch and solvent resistance, and hardness evaluations are subjective in nature and limited in real-life applications. NIR measurement is objective as it provides a method of measuring the actual conversion or disappearance of functional groups during the advancement of curing reactions. The changes in absorbance with time and temperature provide the actual staged degree of epoxy curing.

Sample handling is a critical variable in the spectroscopic testing of these matrices. Over the period of the cure, the samples can be converted from the liquid state to a gel state to the solid state. NIR enables the resin to be scanned as is without the need to dissolve it in solvents. Disposable, glass sampling windows can be used. If the sample is nonhomogeneous, large areas of the sample can be averaged to represent an average state of cure advancement. In addition, the N—H and O—H bands can be resolved in the overtone and combination band region whereas the fundamental absorbances of these bands are often overlapping in the mid-IR region.

Danneberg [91] analyzed epoxy resins for epoxide equivalent by NIR. One milliliter of molten resin was pressed between glass plates using 1/16 in. Teflon spacers. After the resin cooled, the glass plates were removed, leaving epoxy “buttons” that were inserted directly into the spectrophotometer. Band assignments were made for the terminal epoxy overtone at 1159 nm and the combination at 2208 nm. These bands are due to C—H stretching and C—H deformation and are shifted away from other C—H bands due to the electronegativity of the strained oxirane structure.

For epoxy equivalent measurement, Danneberg used an *isobestic* point to overcome problems created by hydrogen bonding. Over the period of the cure, the hydrogen bonding changes dramatically. The OH group when hydrogen bonding occurs at longer wavelengths than the free OH. He observed that the increase in the hydroxy peak at 1428 nm stops after 1 h. Then, the OH broadens and a shoulder appears at 1470 nm. Upon solidification, a different type of H bonding appears to replace the previous type. He defined an isobestic point as a wavelength that meets the following three conditions:

1. Conversion of one absorbing compound, X, into another one, Y, in such a way that the concentration of X + Y, expressed in equivalents per volume unit, remains constant.
2. Absorption bands of X and Y must overlap.
3. Beer's law must be obeyed for both compounds at the isobestic point. The sum of X + Y must be a constant value regardless of the single absorption values of X and Y.

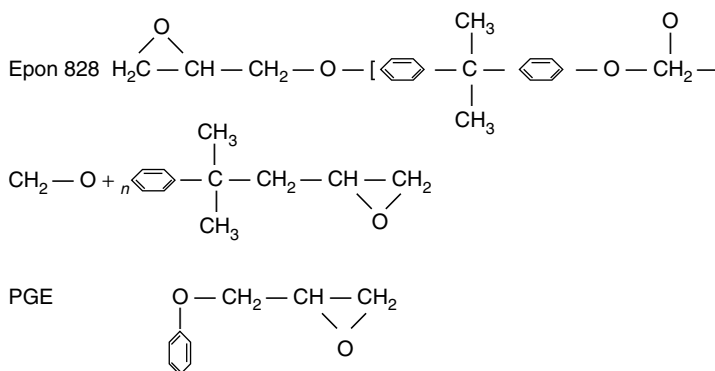
The use of the isobestic point for OH eliminates the problem of band shifting due to hydrogen bonding [91].

Goddu and Delker [92] analyzed terminal epoxides with accuracy and precision of $\pm 2\%$. Weyer [20] measured the intact epoxide group of an allyl glycidyl ether grafted to polypropylene with the 2200-nm band. Samples were presented to the analyzer as 60-mil plaques.

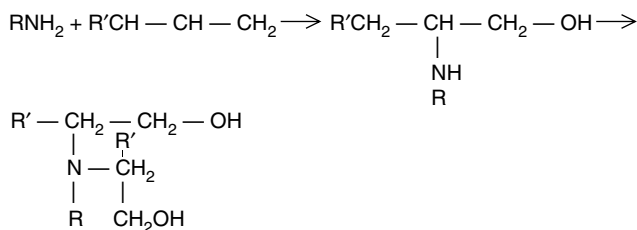
Several epoxy/amine crosslinking reactions were studied by monitoring the disappearance of reactants and the formation of products [93]. The organic functional groups that change as crosslinking takes place are:

- Disappearance of the epoxy reactant at 2208 nm.
- Disappearance of the NH of the primary amine crosslinking agent at approximately 1500 and 2000 nm.
- Formation of the NH of a secondary amine at approximately 1500 nm.
- Eventual disappearance of the secondary amine band at 1500 nm as the tertiary amine is formed.
- Formation of the O—H as the oxirane ring is opened at approximately 1400 nm.

The epoxy resins used in the study were (a) a resin prepared from bisphenol A and epichlorohydrin (Epon 828) and (b) a phenyl glycidyl ether (PGE) resin. Crosslinking agents were hexamethylenediamine (HMDA) and ADH. The chemical formulas are shown as follows:



The basic reaction mechanism of the crosslinking reaction is as follows:



Reactants combined in 1:1 and 1:2 ratios are reacted for various amounts of time. Table 27.6 compares the change in the primary NH combination band at approximately 2000 nm, the primary/secondary NH overtone band at approximately 1500 nm, and the epoxy combination band at 2208 nm under the various experimental conditions. The CH combination band due to the epoxy ring at 2208 nm disappears even under mild curing conditions with HMDA. The 2208 nm peak is still present in the ADH/PGE system after 0.5 h, but disappears after 1 h. The combination band in the 2028 to 2040 nm region is due to the aliphatic primary amine group in the curing agents. Neither the secondary nor tertiary amine exhibits a combination band in this region. The 2040 nm band in the ADH/PGE 1:2 systems is strong after 0.5 h, but decreases in absorbance after 1 h and even more so after 2 h. This denotes the decrease in concentration of primary amine as the curing reaction takes place. In the ADH/PGE; 1:4 system, the 2040 nm band is of medium intensity after 0.5 h, but disappears after 1 h. The disappearance of the band denotes the disappearance of primary amine in the curing system.

The first overtone of the aliphatic NH symmetrical stretch occurs in the 1500 nm region. Unlike the combination band, both primary and secondary amines absorb in this region. Interference from the first overtone of the OH stretching vibrations at 1400 nm is easily avoided with the high resolution available from NIR instruments.

The HMDA/PGE 1:2 system has both the overtone and combination bands after 0.5 h, but neither after 1 h. The disappearance of both bands after 1 h denotes that there is a negligible amount of primary and secondary amine in the curing system after 1 h. In the ADH/PGE 1:2 system, the primary amine absorbance bands steadily decrease upon curing. However, the band at 1556 nm decreases after 0.5 h, and then increases after 1 h. The initial decrease in absorbance is due to the decreasing concentration

TABLE 27.6

Comparison of Primary N—H Combination Band at ~2000 nm, Primary/Secondary Overtone Band at ~1500 nm, and Epoxy Combination Band at 2208 nm

Reactants	Conditions	Wavelengths	Comments
HMDA + EPON	1:1/4 h	1540	Strong, sharp
		2028	Strong, sharp
		2208	Not present
HMDA + PGE	1:2/0.5 h	1544	Strong, sharp
		2028	Strong, sharp
		2208	Not present
HMDA + PGE	1:2/1 h	1544	Not present
		2028	Not present
		2208	Not present
ADH + PGE	1:2/0.5 h	1556	Strong, sharp
		2040	Strong, sharp
		2208	Weak, sharp
ADH + PGE	1:2/1 h	1556	Weak, sharp
		2040	Weak, sharp
		2208	Not present
ADH + PGE	1:2/2 h	1556	Strong, sharp
		2040	Weak, sharp
		2208	Not present
ADH + PGE	1:4/0.5 h	1552	Strong, sharp
		2040	Medium, sharp
		2208	Weak, sharp
ADH + PGE	1:4/1 h	1548	Medium, sharp
		2040	Not present
		2208	Not present
ADH + PGE	1:4/2 h	1552	Medium, sharp
		2040	Not present
		2208	Not present

of primary amine. The subsequent increase in absorbance is due to the formation of secondary amines. The same phenomenon is observed in the ADH/PGE 1:4 system. To further explore the reactions of the amine curing agent in the epoxy-amine cure, a precision scanning system was set up to scan continuously. A 1:2 ratio of bisphenyldiglycidyl ether and (dimethyl) aminoethylphenol (DMP) were mixed together on a glass slide. A cover slide was pressed on top of the mixture. The sample cell was placed on a flat ceramic sample holder. The material was scanned from 1100 to 2500 nm in a continuous mode. The initial reading was measured at 1.2 min and the final reading at 188.2 min. A calibration was developed to predict time of reaction.

The calibration results are reported in the following table:

Epoxy/amine cure	
Range, min	1.2–188.2
SEE	0.82
r^2	0.9999
F value	13,833

The data were plotted in a log–log mode. The $\log(1/R)$ absorbance at 2080 nm was plotted on the y -axis and the $\log(1/R)$ absorbance at 1506 nm was plotted on the x -axis. As the reaction takes

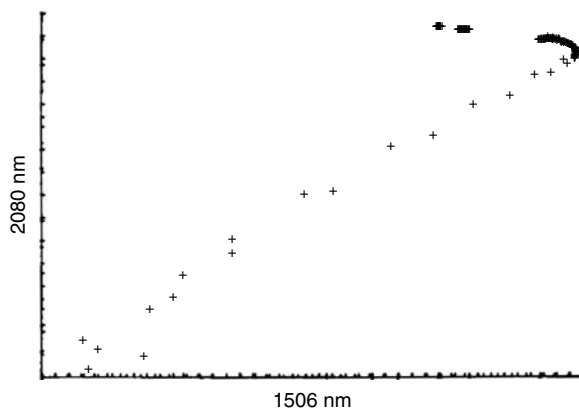


FIGURE 27.10 Conversion of a primary amine to a secondary amine to a tertiary amine during an epoxyamine cure. The absorbance at 1506 vs. 2080 nm was plotted for 100 spectra.

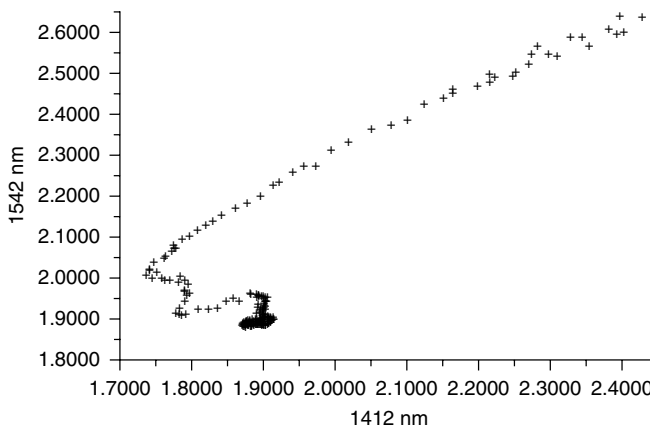


FIGURE 27.11 Conversion of a primary amine to a secondary amine to a tertiary amine during an epoxyamine cure. The absorbance at 1412 vs. 1542 nm was plotted for 100 spectra. Additional information about the reaction kinetics is available because the data were obtained using AOTS technology, which enables ultrafast scanning.

place, the absorbance of the 1506 nm band relative to the 2080 nm band increases steadily. It then reaches an inflection point and begins to decrease relative to the 2080 nm (Figure 27.10).

This boomerang-shaped pattern of absorbance may be attributed to an initial increase in absorbance of the NH overtone band as the primary amine is converted to a secondary amine and then followed by a decrease in the absorbance of the NH overtone band as the secondary amine is converted to a tertiary amine.

A similar boomerang-shaped pattern occurs in the reaction of bisphenyldiglycidyl ether and *N*-aminoethylpiperazine (AEP). A 1:2 ratio was mixed and placed in a glass test tube placed between two single fibers. An AOTS NIR analyzer was used for this experiment. The samples were scanned from 800 to 1700 nm. For this experiment, the system was set to scan at the rate of one complete scan per second.

The $\log(1/R)$ data at 1542 and 1412 nm was plotted on a log-log chart. An initial decrease of absorbance of the 1542 nm peak is seen. Then a slight increase in absorbance occurs. This is followed by a final decrease in absorbance that levels off after about 5 min (Figure 27.11) The spectroscopy of this reaction appears to support the conversion from primary amine to secondary amine to tertiary amine species over the course of the curing reaction.

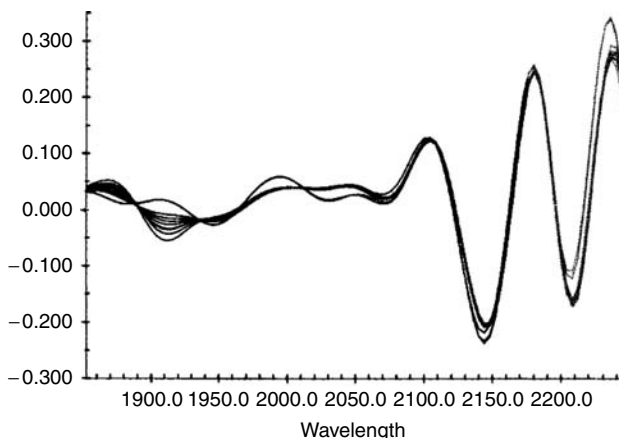


FIGURE 27.12 Decrease of the C—H of the epoxy group at 2208 nm and the appearance of the O—H group in the 1900-nm region for an epoxycarboxylic acid cure system.

An epoxy/carboxylic acid anhydride (1:1) cure was studied in the same manner. The reaction mechanism begins with a ring-opening step to generate a half ester. Then, the carbonyl group reacts with epoxy to form a hydroxydiester. The spectrum obtained by continuously scanning the reaction show the disappearance of the epoxy band at 2208 nm and the appearance of the OH group at 1980 nm (Figure 27.12).

Solids content was determined in epoxy resin using NIR. The sample was poured into a 0.5-mm viscous cell and analyzed in a filter NIR analyzer. Samples ranged from 76 to 86%, with an SEE of 0.35% and an r^2 of 0.994. Samples of different colors were combined in one equation using indicator variables [94].

Resin formation is usually broken down to three stages: A, B, and C. The A-stage resin is usually a thermoplastic that is insoluble in organic solvents. This material is typically dumped from a kettle, cooled, and then ground to a fine powder.

Additives including fillers and colorants are added to the A-stage resin. Upon heating and further reaction, a B-stage resin is formed. B-stage resins are barely soluble in organic solvents but are fusible under heat and pressure. The final step is the fabrication of B-stage resins into products or parts during the C stage. C-stage resins are infusible, crosslinked materials.

It is advantageous to analyze the A- and B-stage resins for key parameters to avoid fabrication of inferior products in the C stage. Simultaneous analysis of resin content and flow content in B-stage resins was achieved using NIR.

Glass/epoxy B-stage resins were analyzed in a standard closed cup. Resin content refers to the amount of resin in the composite. Flow content is an estimate of the degree of cure. Results are summarized in the following table:

	Resin content	Flow content
Range, %	33–42	9–17
Number of samples	10	10
SEE	0.8	1.1
r^2	0.951	0.919
Wavelengths	1445	1818
	1818	2139
	1940	2180

TABLE 27.7
Calibration Results with Standard Cell and EDAPT-1

	Standard cell		EDAPT-1	
	SEE	MCC	SEE	MCC
Percent resin	0.79	0.9511	0.27	0.9933
Flow content	1.1	0.919	0.17	0.998

The sample set was run with an EDAPT-1 probe [56], a referenced fiber-optic probe with an integrating sphere located at the tip of the fiber bundles. The SEE and the r^2 improved on both percent resin and flow content. Table 27.7 compares the two methods of sample preparation.

The improvement is attributed to the fact that the sample is directly in contact with the sphere of the fiber-optic probe whereas in the instrument, there is a finite gap and an additional sample window between the sample and the integrating sphere [95]. The wavelengths chosen for the equations generated using the EDAPT-1 probe were 2206, 2346, and 2444 for resin; and 1394, 1632, and 1870 for flow content.

NIR has been successfully applied to a wide variety of B-stage composite materials, listed in the subsequent sections of this chapter.

The technique can measure large areas of sample and average nonhomogeneous sample readings. In addition, the EDAPT-1 fiber probe can be placed directly on the sample, enabling nondestructive analysis. Resin content and flow content (degree of cure) can be simultaneously determined.

Epoxy content was analyzed on white and black prepreg sheets using the EDAPT-1 probe. Each sheet was scanned at two orientations and averaged. Wavelengths chosen were 1254, 1268, and 2164 nm. The range of epoxy resin was 30 to 33% with an SEE of 0.2 and an r^2 of 0.991. Two distinct calibrations were required.

NIR spectroscopy was used to quantify the cure reaction of MDEA-epoxy resins ($E/A = 1.4$) carried out at 72 and 160°C. A new assignment for the secondary amine at 6580 cm^{-1} was proposed, supported by a synthesized model compound. Two different spectral treatments were proposed. One was based on normalization at 4610 to 4620 cm^{-1} while the another was based on the subtraction of the normalized spectrum of a cured sample. Quantitative results were obtained from absorbance measurements in the combination region, while spectral decompositions and area measurements were necessary in the overtone region. The resulting complementary information allowed calculation of conversion rates of epoxide and amine I, and concentrations of amine II, amine III, hydroxyl groups, and ether links [96].

Simultaneous dielectric and NIR measurements were performed in “real-time” to follow polymerization reactions on blends of a diglycidyl ether of bisphenol A epoxy resin with 4,4'-diaminodiphenylmethane hardener and different amounts of poly(methyl methacrylate) as modifier. The effect of the curing temperature and modifier amount on the polymerization reactions were studied. Epoxy and amine conversions were followed by NIR, while changes in molecular mobility in the reaction mixture were analyzed by dielectric relaxation spectroscopy (DRS). Evolutions of ionic conductivity and relaxation were analyzed and vitrification times were obtained. The relaxation behavior was analyzed through curing in the frequency domain, being the change of the main relaxation indicative of the cure reaction advancement. DRS data were also presented as complex impedance. Vitrification times, obtained by dielectrometry were compared with those obtained by rheological measurements and gelation times obtained by NIR were compared with those obtained by solvent extraction [97].

The kinetics of network formation of an amine-terminated linear or three-arm star poly(propylene oxide) and a bifunctional epoxy prepolymer were evaluated by NIR spectroscopy, and the dynamics

were investigated by broad-band DRS and dynamic mechanical spectroscopy (DMS). The dynamics of networks containing linear and star chains were similar but not identical [98].

The microwave and thermal cure processes for the epoxy-amine systems (epoxy resin diglycidyl ether of bisphenol A, DGEBA) with 4,4'-diaminodiphenyl sulfone (DDS) and 4,4'-diaminodiphenyl methane (DDM) were investigated for 1:1 stoichiometries by using fiber-optic FT-NIR spectroscopy. The kinetic rate parameters for the consumption of amines were determined by a least squares curve fit to a model for epoxy/amine cure. The activation energies for the polymerization of the DGEBA/DDS system were determined for both cure processes and found to be 66 and 69 kJ/mol for the microwave and thermal cure processes, respectively. No evidence was found for any specific effect of the microwave radiation on the rate parameters, and the systems were both found to be characterized by a negative substitution effect [99].

The isothermal photopolymerization kinetics of a dimethacrylate and the thermal cure of an epoxy/anhydride mixture were investigated separately and within the 50:50 dimethacrylate/epoxy interpenetrating polymer networks (IPN) using DSC and NIR. This combination of cross-linkable resins permits the partial or complete cure of each component independent of each other. The effect of changing the order of cure of the dimethacrylate or the epoxy within the IPN on the polymerization kinetics of each group was studied using NIR. The conversion of each species was found to be very dependent on the cure order [46]. Previous investigations were done on a model vinyl ester (VER, composed of bisphenol A diglycidyl ether dimethacrylate, bisGMA, in 30% wt. styrene) cured with azobis(isobutyronitrile) (AIBN) and an epoxy monomer (diglycidyl ether of bisphenol A, DGEBA) cured either with 1-methylimidazole (1-MeI) or with a stoichiometric quantity of 4,4'-diaminodiphenylmethane (DDM). NIR studies of other monomer combinations showed that the vinyl ester resin component cured more quickly than the epoxy component. The rate of vinyl conversion was slowed due to dilution by the epoxy system and retardation of the radical polymerization by the amines. The rate of epoxy conversion in the mixture was slower than in the neat epoxy, indicating that the presence of the vinyl ester component also had a dilutional effect. The rate of epoxy conversion was faster when catalyzed by hydroxy groups in bisphenol A diglycidyl ether dimethacrylate than in the neat epoxy. The rheology and gel times indicated that the gelation of the VER component determined the overall gel behavior of the IPN. The gel times correlated well with the vinyl NIR conversion data [100].

A commercial grade of a tetrafunctional epoxy resin (TGDDM) was modified by the addition of a perfluoro-ether oligomer to improve the toughness. The cure characteristics of a typical composition of this mixture were investigated by *in situ* FT-NIR spectroscopy. The presence of the, fluoro-oligomer induced a two-fold reduction of the initial curing rate [101].

FT-NIR was used to investigate the molecular interactions occurring between absorbed water molecules and networks based on a tetrafunctional epoxy resins. One formulation contained DDS as a hardener, and the other was a modified resin containing 4,4'-bismaleimide-diphenylmethane (BMI) as a coreactive monomer. NIR spectroscopy analysis confirmed the existence of mobile water localized in network defects (microvoids) that did not interact with the networks and water molecules bound to the networks through hydrogen-bonding interactions. In the BMI-containing system, the fraction of bound water decreased significantly with respect to the unmodified epoxy resin. This was a relevant result because the bound water was primarily responsible for the plasticization of the network and for the consequent worsening of mechanical performance. Water diffusion was investigated with gravimetric sorption measurements and time-resolved FT-NIR spectroscopy measurements. These showed that the presence of BMI decreased the water uptake at equilibrium, enhanced the diffusivity, and reduced the activation energy for diffusion. A dual-mode model for diffusion was found to be suitable for accurately describing the mass-transport process and allowed estimation the ratio of free and bound water, which was in good agreement with that obtained from the spectroscopic analysis [102].

27.3.14 PHENOLIC RESINS

The amount of hexamethylenetetramine in A-stage phenolic resins was determined using NIR. The resin material was sampled in a fine powder form, using the standard closed cup. Spectral characteristics of the methyl and amine groups of the crosslinking agent are distinct as well as hydroxyl groups and aromatic CH groups of the phenolic resin.

The range of hexamethylenetetramine was 0 to 1% with an SEE of 0.4% and an r^2 of 0.992.

The wavelengths chosen were 1722, 1786, and 1828 nm.

Paper/phenolic and linen/phenolic B-stage resins were analyzed using a precision scanning NIR analyzer. The results are summarized in the following tables:

Paper/phenolic	53–56	I–II
Number of samples	10	10
r^2	0.903	0.953
SEE	0.5	1.0
Wavelengths	1394	1840
	1674	2040
	1702	2336

27.3.15 SILICONES

Silicone lubricants are sprayed on textile fibers to facilitate processing. The application of the lubricant is frequently measured by NIR techniques. A linear trend was observed between percent finish on nylon and absorbance. The Si—OH band and the Si—C—H band are shifted relative to the C—OH band and the C—C—H band due to lesser electronegativity of Si as compared to the C. Ghosh et al. [103] report the measurement of lubricants on nylon materials.

Silicone resins are frequently used to coat materials to provide desired characteristics. The measurement of silicone films on paper products has also been performed. A silicone curing reaction on a paper substrate was studied using NIR. The degree of cure could be followed by the disappearance of the Si—OH bands and the appearance of Si—C—H bands.

NIR and MIR reflectance spectroscopy were used to characterize Whatman No. 1 filter paper coated with solutions of poly(hydrogenmethylsiloxane) and dichlorodimethylsilane. The paper was coated to improve the extraction of polycyclic aromatic hydrocarbons (PAHs) from water [104].

Discriminant analysis was used to group coating quality for silicone-coated polyester film. Results are preliminary but do indicate that NIR can distinguish between “good” and “bad” coating. Wavelengths chosen were 1722, 1940, and 2272 nm. Mahalanobis distances between groups are reported in the following table:

	Good	Bad	Borderline
1722	8.7	106	15.6
1940		2.0	7.8
2272			6

Glass/silicone B-stage resins were analyzed using NIR. Calibration results are summarized in the following table:

	Resin content	Flow content
Range, %	30–33	0.3–2.7
Number of samples	12	12
r^2	0.902	0.716
SEE	0.4	0.4
Wavelengths	1680	2180
	2250	2250
	2336	2336

The polymerization of 1H, 1H, 2H, 2H-perfluorooctyltriethoxysilane in ethanol, catalyzed by 0.5, 1.0 and 2.0 M HCl . H₂O, was examined using time-resolved NIR and 2D NIR correlation techniques. The time-resolved NIR results have demonstrated that the growth of polymeric aggregates prior to the phase separation proceeds in a two-step process, which depends upon the HCl concentration. Furthermore, it has been found that the 2D NIR correlation data provided direct information on the sequence of the reaction steps and interaction of participating components through hydrogen bonding in the growth process [105].

27.3.16 FORMALDEHYDE

Glass/melamine/formaldehyde B-stage resins have been studied using NIR. Samples were scanned directly with multiple point averaging. The results are summarized in the following table:

	Resin content	Flow content
Range, %	35–39	8–18
Number of samples	10	10
r^2	0.945	0.968
SEE	0.4	0.9
Wavelengths	1445	1632
	1790	2080
	1982	2276

An FT-NIR method was developed for the fast and reliable assessment of the pathway(s) followed during formaldehyde-based resin synthesis at laboratory and industrial scale. No sample manipulation was necessary and the complete evaluation was performed on- or off-line in less than 1 min. The method was found to provide a fast and consistent way of directly monitoring the effects of resin formulation changes when evaluating new procedures at the laboratory scale. Deviations from the standard synthetic pathway were detected on-line and in time to allow taking the necessary steps to return to the desired pathway. Furthermore, NIR methodologies were developed to identify and check the conformity of raw materials and final products including urea, UFC solutions, and laminated paper produced by impregnation with formaldehyde-based resins. This could be applied to laminated paper production where the reproducibility and the effects of storage were both questionable and difficult to assess [106].

FT-NIR spectroscopy via optical fibers was used to monitor the formation of urea-formaldehyde resins *in situ*. The effects of pH and temperature on the structure were studied. GPC, NMR and Raman measurements were used to elucidate the structural characteristics. Resins produced at high

temperatures and acidic pH values exhibited higher degrees of condensation, presumably because of the development of more cross-linked structures [107].

27.4 CONCLUSIONS

Rapid, nondestructive analysis of a wide variety of chemical and physical properties of polymers is achievable using NIR analysis. Raw materials can be quickly tested for identity and quality conformance. Batch-to-batch, lot-to-lot, and drum-to-drum consistency can be determined. Spectral patterns of materials can be archived for future comparison with new materials.

Quality assurance and quality control of polymeric materials is greatly facilitated by the ease of sampling, the reproducibility of results, and the speed of testing.

The use of NIR in polymer process applications has grown tremendously. Results are obtained online in seconds, providing a high measuring frequency, which makes NIR an ideal technique for statistical process analysis. Real-time testing also enables detection of malfunction in time for process corrections to be implemented. Anticipatory process control is especially important for thermosetting materials that, once reacted, are irreversibly formed.

Clearly, the implementation of NIR in polymer manufacturing processes results in improved product quality and overall process efficiency.

REFERENCES

1. Staudinger, H., *Ber. Bunsenges. Phys. Chem.*, 1920. **53**: 1073.
2. Staudinger, H. and J. Fritsch, *Helv. Chim. Acta*, 1922. **5**: 785.
3. Staudinger, H., *Ber. Bunsenges. Phys. Chem.*, 1926. **59**: 3019.
4. Staudinger, H., K. Frey, and W. Starck, *Ber. Bunsenges. Phys. Chem.*, 1927. **66**: 1782.
5. Stane, G.A., *Development of Modern Polymer Theory*. Chemtech, 1984. **August**: p. 492.
6. Hausdorff, H.H., Analysis of Polymers by Infrared Spectroscopy, in *Pittsburgh Conference on Analytical and Applied Spectroscopy*. 1951. Pittsburgh, PA.
7. Kagarese, R.E. and L.A. Weinberger, *Infrared Spectra of Plastics and Resins*. U.S. Department of Commerce, OTS Bull., 1954. **PB 111438**.
8. Herschel, W., *Phil. Trans.*, 1800. **90**: 225.
9. Stark, E. and K. Luchter, *Appl. Spectrosc. Rev.*, 1986. **22**: 4.
10. Taylor, E.S., *Spectroscopy*, 1989. **4**: 10.
11. Goddu, R.F., Near-Infrared Spectrophotometry, in *Advances In Analytical Chemistry and Instrumentation*, C.N. Reilley (Ed.), 1960. Interscience. pp. 347–424.
12. Kaye, W., Near-Infrared Spectroscopy: Part I. Spectral Identification and Analytical Applications. *Spectrochimica Acta*, 1954. **6**: 257–287.
13. Wheeler, O.H., *Near Infrared Spectra of Organic Compounds*. *Chem. Rev.*, 1959. **59**: 629–666.
14. Glatt, L. and J.W. Ellis, Near Infrared Pleochroism II. The 0.3-2.5 Region of Some Linear Polymers. *J. Chem. Phys.*, 1951. **19**: 449–457.
15. Miller, R.G.J. and H.A. Willis, Quantitative Analysis in the 2-micron Region Applied to Synthetic Polymers. *J. Appl. Chem.*, 1956. **6**: 385–391.
16. Crandall, E.W. and A.N. Jagtap, The Near-Infrared Spectra of Polymers. *J. Appl. Polym. Sci.*, 1977. **21**: 449–454.
17. Griffith, D., Mid Infrared and Near Infrared Reflectance Analysis, in *Second Annual Symposium on Near Infrared Reflectance Analysis*. 1982. Tarrytown, NY: Technicon Industrial Systems.
18. Siesler, H., Mid and NIR Infrared Spectroscopy of Polymers. Recent Developments in Molecular Spectroscopy, in *10th National Conference on Molecular Spectroscopy*, 1988. Bulgaria.
19. Davies, A.M.C., et al., Rapid Analysis of Packaging Laminates by Near-Infrared Spectroscopy. *Analyst*, 1985. **110**: 643.
20. Weyer, L.G., Utilizing Zero Crossover Points in the Near Infrared Reflectance Analysis of Ind. *J. Appl. Polym. Sci.*, 1986. **31**: 2417–2431.
21. Ciurczak, E., *Spectroscopy*, 1986. **1**: 7.

22. Hall, J.E., et al., *J. Coat. Technol.*, 1988. **60**: 49–61.
23. Ghosh, S. and J.E. Rodgers, Determining Heat Set Temperature by Near-Infrared Reflectance Spectroscopy. *Text. Res. J.*, 1985. **55**: 556–560.
24. *Operation Manual for Sesame Software*. 1998, Bran + Luebbe: Norderstedt, Germany.
25. Howell, H. and J. Davis, Qualitative Identification of Polymeric Materials Using Near IR Spectroscopy, in *Am. Chem. Soc., PMSE Division Spring Meeting*. 1991. New York, NY.
26. Eilert, A., et al., Near Infrared Analysis of Polymers, in *Pittsburgh Conference*. 1999. Orlando, FL.
27. Rohe, T. and H. Kull, *Plastics Identification-Fast, Reliable, Flexible. Bulletin D-76327*, Fraunhofer Institute Chemische Technologie: Pfinztal (Berghausen), Germany.
28. Camacho, W. and S. Karlsson, NIR, DSC, and FTIR as Quantitative Methods for Compositional Analysis of Blends of Polymers Obtained from Recycled Mixed Plastic Waste. *Polym. Eng. Sci.*, 2001. **41**: 1626–1635.
29. van den Broek, W., et al., Plastic Identification by Remote Sensing Spectroscopic NIR Imaging Using Kernel Partial Least Squares (KPLS). *Chemom. Intell. Lab. Syst.*, 1996. **35**: 187–197.
30. van den Broek, W., et al., Identification of Plastics among Nonplastics in Mixed Waste by Remote Sensing Near-Infrared Imaging Spectroscopy. 1. Image Improvement and Analysis by Singular Value Decomposition. *Anal. Chem.*, 1995. **67**: 3753.
31. Wienke, D., et al., Near Infra-red Imaging Spectroscopy (NIRIS) and Image Rank Analysis for Remote Identification of Plastics in Mixed Waste. *Fresenius' J. Anal. Chem.*, 1996. **354**: 823–828.
32. Wienke, D., W. van den Broek, and L. Buydens, Identification of Plastics among Non-plastics in Mixed Waste by Remote Sensing Near Infrared Imaging Spectroscopy. 2. Multivariate Image Rank Analysis for Rapid Classification. *Anal. Chem.*, 1995. **67**: 3753–3759.
33. Feldhoff, R., et al., On-line Post Consumer Package Identification by NIR Spectroscopy Combined with a Fuzzy ARTMAP Classifier in an Industrial Environment. *Appl. Spectrosc.*, 1997. **51**: 362–368.
34. Wienke, D., et al., Comparison of an Adaptive Resonance Theory Based on Neural Network (ART-2a) against other Classifiers for Rapid Sorting of Post Consumer Plastics by Remote Near Infrared Spectroscopic Sensing Using an InGaAs Diode Array. *Anal. Chim. Acta*, 1995. **317**: 1–16.
35. Operation Manual for Discriminant Analysis Software, in *Technical Publication No. DSA-008-01*. 1989, Bran and Luebbe Analyzing Technologies: Elmsford, NY.
36. Siesler, H.W., Near Infrared Spectroscopy for Polymer Reaction and Process Control: Synthesis, Processing and Recycling. *NIR News*, 1995. **6**: 3–6.
37. Williams, S. and H.L. McPeters, NIR Spectroscopy in Real Time Analysis of Polymer Melt Processes, in *Making Light Work: Advances in Near-Infrared Spectroscopy*, I. Murray and I. Cowe (Eds.), 1991. VCH: Aberdeen, Scotland. pp. 544–553.
38. Brimmer, P.J., S.L. Monfre, and F.A. DeThomas, Real-Time Monitoring for the Production of Polyurethane, in *Making Light Work: Adv. Near Infrared Spectrosc.*, Int. Conf. on NIR, I. Murray and I.C. Cowe (Eds.), 1991. Aberdeen, Scotland. pp. 554–565.
39. Lammers, H., M.P.B. Vanuum, and J.P. Dekleijn, Process Analysis — Properties of Poly(Ethylene-Terephthalate) Measured by Near-Infrared Spectroscopy. 2. In-Line Analysis of Poly(Ethylene-Terephthalate) Melt. *Macromol. Chem. Phys.*, 1995. **196**: 2029–2034.
40. Van der Elst, R.H., The Implementation of On-Line NIR (AOTS) for Control of a Polymerization Process in a Batch Reactor, in *KVIV-BIRA Symposium*. 1997. Antwerp.
41. Nagata, T., M. Tanigaki, and M. Ohshima, In-Line NIR sensing of CO₂ concentration in polymeric foaming extrusion process. *J. Cell. Plastics*, 2002. **38**: 11–30.
42. Masaki, K., et al., Fourier-Transform Near-Infrared and Electron-Spin-Resonance Studies on the Cross-Linking Reaction of Liquid Carboxylated Poly(Acrylonitrile-Co-Butadiene) Rubber with Dicumyl Peroxide in Dioxane. *J. Appl. Polym. Sci.*, 2003. **89**: 2095–2101.
43. Tosi, C. and A. Pinto, Near-Infrared Spectroscopy of Hydrocarbon Functional Groups. *Spectrochim. Acta*, 1972. **28A**: 585–587.
44. Macho, S., et al., Outlier Detection in the Ethylene Content Determination in Propylene Copolymer by Near-Infrared Spectroscopy and Multivariate Calibration. *Appl. Spectrosc.*, 2001. **55**: 1532–1536.
45. Furukawa, T., et al., Discrimination of Various Poly(Propylene) Copolymers and Prediction of Their Ethylene Content by Near-Infrared and Raman Spectroscopy in Combination with Chemometric Methods. *J. Appl. Polym. Sci.*, 2003. **87**: 616–625.
46. Tosi, C., *Makromol. Chem.*, 1968. **112**: 303.

47. Bly, R.M., P.E. Kiener, and B.A. Fries, Near Infrared Method for Analysis of Block and Random Ethylene–Propylene Copolymers. *Anal. Chem.*, 1966. **38**: 217–220.
48. Takeuchi, T.S.T. and Y. Sugimura, Near-infrared Spectrophotometric Analysis of Ethylene-Propylene Copolymers. *Anal. Chem.*, 1969. **41**: 184–186.
49. Miller, C.E., Use of Near-Infrared Spectroscopy to Determine the Composition of High-Density Low-Density Polyethylene Blend Films. *Appl. Spectrosc.*, 1993. **47**: 222–228.
50. Watari, M., et al., On-Line Monitoring of the Density of Linear Low-Density Polyethylene in a Real Plant by Near-Infrared Spectroscopy and Chemometrics. *Appl. Spectrosc.*, 2004. **58**: 248–255.
51. Camacho, W. and S. Karlsson, Simultaneous Determination of Molecular Weight and Crystallinity of Recycled HDPE by Infrared Spectroscopy and Multivariate Calibration. *J. Appl. Polym. Sci.*, 2002. **85**: 321–327.
52. Weyer, L.G., Near-Infrared Spectroscopy of Organic Substances. *Appl. Spectrosc. Rev.*, 1985. **21**: Jan-43.
53. Camacho, W. and S. Karlsson, Quantification of Antioxidants in Polyethylene by Near Infrared (NIR) Analysis and Partial Least Squares (PLS) Regression. *Int. J. Polym. Analysis Charact.*, 2002. **7**: 41–51.
54. Lachenal, G., I. Stevenson, and N. Celette, Near-Infrared Transmittance Spectroscopy for Radiochemical Ageing of EPDM. *Analyst*, 2001. **126**: 2201–2206.
55. Lachenal, G. and I. Stevenson, Applicability of Fourier Transform Near Infrared Spectroscopy and Principal Components Analysis for Studying Elastomers after Radiochemical Ageing. *J. Near Infrared Spectrosc.*, 2002. **10**: 279–287.
56. Operation Manual for the EDAPT-1, in *Technical Publication No: DSM-0007-01*. 1989, Bran and Luebbe Analyzing Technologies: Elmsford, NY.
57. Laasonen, M., et al., Near Infrared Reflectance Spectroscopy for the Fast Identification of PVC-Based Films. *Analyst*, 2001. **126**: 1122–1128.
58. Abd El-Kader, K.A.M., et al., Effect of the Molecular Weights on the Optical and Mechanical Properties of Poly(Vinyl Alcohol) Films. *Polym. Test.*, 2002. **21**: 847–850.
59. Weyer, L.G., The Use of Derivative Nodes in Near-Infrared Spectroscopy, in *Analytical Applications of Spectroscopy (Incorporating the First International Near Infrared Spectroscopy Conference)*. 1988. Norwich, UK: Royal Society of Chemistry.
60. Lamere, L., Applications of Near Infrared Reflectance Analysis in Warp Sizing: Determination of Size Add-On, in *TQCA Meeting*. 1990. Air Products and Chemicals, Inc.
61. Kirkbright, G.F. and K.R. Menon, The Determination of Combined Vinyl Acetate in Polyvinyl Chloride-Polyvinyl Acetate Copolymer by PAS and NIR. *Anal. Chim. Acta*, 1982. **136**: 373.
62. Furukawa, T., et al., On-Line Monitoring of Melt-Extrusion Transesterification of Ethylene Vinylacetate Copolymers by Near Infrared Spectroscopy and Chemometrics. *J. Near Infrared Spectrosc.*, 2002. **10**: 195–202.
63. Ozaki, Y., et al., Self-Modeling Curve Resolution Analysis of On-Line Vibrational Spectra of Polymerisation and Transesterification. *Macromol. Symp.*, 2002. **184**: 229–247.
64. Dzwinczyk, M. and H. Mark, Method for Daily Check of Wavelength Accuracy in Near-Infrared Reflectance Spectrophotometers. *Spectroscopy (Springfield, Oreg.)*, 1986. **1**: 51.
65. Giammarise, A., Near Infrared Analysis of the Styrene Content of Copolymers with Aliphatic Acrylates and Methacrylates. *Anal. Lett.*, 1969. **2**: 117–121.
66. Greive, W.H. and D.D. Doepkken, The Determination of Acrylonitrile-Butadiene-Styrene Terpolymer in PVC Using Infrared or Near Infrared Spectrophotometry. *Polym. Eng. Sci.*, 1968. **January**: 19–23.
67. Fontoura, J.M.R., et al., Monitoring and Control of Styrene Solution Polymerization Using NIR Spectroscopy. *J. Appl. Polym. Sci.*, 2003. **90**: 1273–1289.
68. Kaur, S., G. Florio, and D. Michalak, Cross-Linking of Sulfonated Styrene-Ethylene/Butylene-Styrene Triblock Polymer via Sulfonamide Linkages. *Polymer*, 2002. **43**: 5163–5167.
69. Ren, Y., T. Murakami, and T. Nishioka, Two-Dimensional Near-Infrared Correlation Spectroscopy Studies of Compatible Polymer Blends: Composition-Dependent Spectral Variations of Blends of Atactic Polystyrene and Poly(2,6-Dimethyl-1,4-Phenylene Ether). *J. Phys. Chem. B*, 2000. **104**: 679–690.
70. Miller, C.E. and T.K. Yin, Near-Infrared Reflectance Analysis of Poly(Octadecyl Methacrylate) Adsorbed on Alumina. *J. Mater. Sci. Lett.*, 1989. **8**: 467–469.

71. Shuhui Qin, N., et al., Polymer Brushes on Single-Walled Carbon Nanotubes by Atom Transfer Radical Polymerization of *n*-Butyl Methacrylate. *J. Amer. Chem. Soc.*, 2004. **126**: 170–176.
72. Cherfi, A. and G. Fevotte, On-Line Conversion Monitoring of the Solution Polymerization of Methyl Methacrylate Using Near-Infrared Spectroscopy. *Macromol. Chem. Phys.*, 2002. **203**: 1188–1193.
73. Cherfi, A., G. Fevotte, and C. Novat, Robust On-Line Measurement of Conversion and Molecular Weight Using NIR Spectroscopy During Solution Polymerization. *J. Appl. Polym. Sci.*, 2002. **85**: 2510–2520.
74. Le, T.T. and D.J.T. Hill, Simultaneous FT-NIR and ESR Analyses to Yield Propagation Rate Coefficients for Polymerization of Methyl Methacrylate Based Monomers. *Polym. Int.*, 2003. **52**: 1694–1700.
75. Vieira, R.A.M., et al., In-Line and in situ Monitoring of Semi-Batch Emulsion Copolymerizations Using Near-Infrared Spectroscopy. *J. Appl. Polym. Sci.*, 2002. **84**: 2670–2682.
76. Gerasimov, T.G. and D.L. Snavely, Vibrational Overtone Spectroscopy of Ethylene Glycol Diacrylate and Ethylene Glycol Dimethacrylate, Monomer and Polymer. *Appl. Spectrosc.*, 2002. **56**: 212–216.
77. Mark, H., Qualitative NIR Analysis Using Mahalanobis Distances. *Anal. Chem.*, 1985. **57**: 1449–1456.
78. Operation Manual for Cascade Software, in *Technical Publication No. DSA-0025-01*. 1989, Bran and Luebbe Analyzing Technologies: Elmsford, NY.
79. Miller, C.E. and B.E. Eichinger, Analysis of Rigid Polyurethane Foams by Near-Infrared Diffuse Reflectance Spectroscopy. *Appl. Spectrosc.*, 1990. **44**: 887–894.
80. Miller, C.E., P.G. Edelman, and B.D. Ratner, Near-Infrared Spectroscopic Analyses of Poly(Ether Urethane Urea) Block Copolymers. Part I: Bulk Composition. *Appl. Spectrosc.*, 1990. **44**: 576–580.
81. Miller, C.E., et al., NIR Analyses of Poly(Ether Urethane Urea) Block Copolymers. Part II: Phase Separation. *Appl. Spectrosc.*, 1990. **44**: 581.
82. Miller, C.E., et al., Determination of Physical Properties of Reaction-Injection-Molded Polyurethanes by NIR-FT-Raman Spectroscopy. *Appl. Spectrosc.*, 1990. **44**: 1297–1300.
83. Dupuy, J., et al., FT-NIR Monitoring of a Scattering Polyurethane Manufactured by Reaction Injection Molding (RIM): Univariate and Multivariate Analysis versus Kinetic Predictions. *Macromol. Symp.*, 2002. **184**: 249–260.
84. Kienle, R.H. and A.G. Hovey, *J. Am. Chem. Soc.*, 1930. **52**: 3636.
85. Roy, R.B. and C. Kradjel, Application of NIRA Techniques for the Determination of Polymer End and Functional Groups. *J. Polym. Sci., Part A: Polym. Chem.*, 1988. **26**: 1733–1742.
86. Taylor, G.S., *Internal Communication*. 1989, Bran and Luebbe Analyzing Technologies: Elmsford, NY.
87. Miller, C.E. and B.E. Eichinger, Determination of Crystallinity and Morphology of Fibrous and Bulk Poly(Ethylene Terephthalate) by Near-Infrared Diffuse Reflectance Spectroscopy. *Appl. Spectrosc.*, 1990. **44**: 496–504.
88. Amari, T. and Y. Ozaki, Real-Time Monitoring of the Oligomerization Reaction of Bis(hydroxyethylterephthalate) by Near-Infrared Spectroscopy and Chemometrics. *Appl. Spectrosc.*, 2002. **56**: 350–356.
89. Amari, T. and Y. Ozaki, Generalized Two-Dimensional Attenuated Total Reflection/Infrared and Near-Infrared Correlation Spectroscopy Studies of Real-Time Monitoring of the Initial Oligomerization of Bis(Hydroxyethyl Terephthalate). *Macromolecules*, 2002. **35**: 8020–8028.
90. Larena, A., et al., Effect of Surface Roughness on the Optical Properties of Multilayer Polymer Films. *Appl. Surface Sci.*, 2002. **187**: 339–346.
91. Danneburgh, H., *SPE Trans.*, 1963. **January**: 78–88.
92. Goddu, R.F. and D.A. Delker, Determination of Terminal Epoxides by Near-Infrared Spectrophotometry. *Anal. Chem.*, 1958. **30**: 2013–2016.
93. Kradjel, C., Paper Number 833. Determination of the State of Epoxy Cure Advancement Using NIRA Procedures, in *The Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy*. 1987. New Orleans, LA.
94. Burns, D.A. and H. Mark., Indicator Variables in NIRA: How to Use Them, in *7th Int. Symposium on Near Infrared Reflectance Analysis NIRA, Technicon*. 1985. Tarrytown, NY.
95. Kemeny, G., *Internal Communication*. 1988, Bran and Luebbe Analyzing Technologies: Elmsford, NY.
96. Billaud, C., R. Legras, and V. Carlier, Quantitative Analysis of Epoxy Resin Cure Reaction: A Study by Near-Infrared Spectroscopy. *Appl. Spectrosc.*, 2002. **56**: 1413–1421.

97. Kortaberria, G., et al., Curing of an Epoxy Resin Modified with Poly(Methylmethacrylate) Monitored by Simultaneous Dielectric/Near Infrared Spectroscopies. *Europ. Polym. J.*, 2004. **40**: 129–136.
98. Mijovic, J., et al., Interplay of Segmental and Normal Mode Dynamics in Polymer Networks Undergoing Chemical Cross-Linking. Epoxy/Amine-Terminated Linear and Star PPO Formulations. *Macromolecules*, 2003. **36**: 4589–4602.
99. Hill, D.J.T., G.A. George, and D.G. Rogers, A Systematic Study of the Microwave and Thermal Cure Kinetics of the Dgeba/Dds and Dgeba/Ddm Epoxy-Amine Resin Systems. *Polymers for Advanced Technologies*, 2002. **13**: 353–362.
100. Dean, K., et al., Near-Infrared and Rheological Investigations of Epoxy-Vinyl Ester Interpenetrating Polymer Networks. *Macromolecules*, 2001. **34**: 6623–6630.
101. Musto, P., et al., Cure Kinetics and Ultimate Properties of a Tetrafunctional Epoxy Resin Toughened by a Perfluoro-Ether Oligomer. *Polymer*, 2001. **42**: 5189–5198.
102. Musto, P., et al., Probing the Molecular Interactions in the Diffusion of Water through Epoxy and Epoxy-Bismaleimide Networks. *J. Polym. Sci. Part B-Polym. Phys.*, 2002. **40**: 922–938.
103. Ghosh, S., et al., *International Proceedings Institute of Textile Technology, Biannual Report*, 1996. Charlottesville, VA.
104. Ackerman, A.H. and R.J. Hurtubise, Methods for Coating Filter Paper for Solid-Phase Microextraction with Luminescence Detection and Characterization of the Coated Filter Paper by Infrared Spectrometry. *Anal. Chim. Acta*, 2002. **474**: 77–89.
105. Izawa, K., et al., Growth Process of Polymer Aggregates Formed by Perfluorooctyltriethoxysilane. Time-Resolved Near-IR and Two-Dimensional Near-IR Correlation Studies. *Colloid and Polymer Science*, 2002. **280**: 380–388.
106. Dessipri, E., et al., Use of FT-NIR Spectroscopy for On-Line Monitoring of Formaldehyde-Based Resin Synthesis. *Europ. Polym. J.*, 2003. **39**: 1533–1540.
107. Minopoulou, E., et al., Use of NIR for Structural Characterization of Urea-Formaldehyde Resins. *Int. J. Adhesion and Adhesives*, 2003. **23**: 473–484.
108. Lu, H., L.G. Lovell, and C.N. Bowman, Exploiting the Heterogeneity of Cross-Linked Photopolymers to Create High-T-G Polymers from Polymerizations Performed at Ambient Conditions. *Macromolecules*, 2001. **34**: 8021–8025.
109. Reis, M.M., et al., Correlation between Polymer Particle Size and in-situ NIR Spectra. *Macromol. Rapid Commun.*, 2003. **24**: 620–624.
110. Sasic, S., Y. Katsumoto, and H. Sato, Applications of Moving Window Two-Dimensional Correlation Spectroscopy to Analysis of Phase Transitions and Spectra Classification. *Anal. Chem.*, 2003. **75**: 4010–4018.
111. Schilli, C., M.G. Lanzendorfer, and A.H.E. Muller, Benzyl and Cumyl Dithiocarbamates as Chain Transfer Agent in the RAFT Polymerization of N-isopropylacrylamide. In situ FT-NIR and MALDI-TOF MS Investigation. *Macromolecules*, 2002. **35**: 6819–6827.
112. Camacho, W., M. Hedenqvist, and S. Karlsson, Near Infrared (NIR) Spectroscopy Compared with Thermogravimetric Analysis as a Tool for On-Line Prediction of Water Diffusion in Polyamide 6,6. *Polym. Int.*, 2002. **51**: 1366–1370.
113. Wu, P. and H.W. Siesler, Fourier Transform NIR Study of Liquid Diffusion Processes in Nylon 11 Films: Comparison of Water with Alcohols. *Chem. Materials*, 2003. **15**: 2752–2756.
114. De Araujo, S.C. and Y. Kawano, Near-Infrared Spectra of Polyamide 6, Poly(Vinyl Chloride), and Polychlorotrifluoroethylene. *J. Appl. Polym. Sci.*, 2002. **85**: 199–208.
115. Hill, D.J.T., et al., The Radical Homopolymerization of n-phenylmaleimide, N-N-Hexylmaleimide and N-Cyclohexylmaleimide in Tetrahydrofuran. *Polymer*, 2001. **42**: 4791–4802.
116. Mitchell, J.A., C.D. Bockman, and A.V. Lee, Determination of Acetyl Content of Cellulose Acetate by Near-Infrared Spectroscopy. *Anal. Chem.*, 1957. **29**: 499–502.
117. De Wit, J.S. and D. Dugger, Routine Testing of Cellulose Esters by Near Infrared Spectroscopy, in *Process Control and Quality*, 1992. **2**: 123.
118. Hilton, C.L., *Anal. Chem.*, 1959. **31**: 1610.
119. Lefevre, R.J.W., G.M. Parkins, and R. Roper, The Spectroscopic Estimation of Molecular Weights of Polyethylene Glycol Dissolved in Benzene. *Aust. J. Chem.*, 1966. **13**: 169.
120. Jones, C. and J.A. Brown, Polyether Polyol Monitoring Using Near-Infrared Process Photometers. *J. Adv. Inst.*, 1983. **38**: 429.

121. Byron, T., Fast OH Number: Determination for Process Control, in *2nd Annual NIRA Symposium, Technicon*. 1982. Tarrytown, NY.
122. Turley, P. and A. Pietrantonio, Rapid Hydroxyl Number Determination by Near Infrared Reflectance Analysis. *J. Cell. Plastics*, 1984. **20**: 274–277.
123. Grob, R.L., et al., Remote On-line Monitoring of Polyol Production Using Near Infrared Spectrophotometry. *Process Cont. Qual.*, 1992. **2**: 225–235.
124. Tan, H., S.T. Sum, and S.D. Brown, Improvement of a Standard-Free Method for Near-Infrared Calibration Transfer. *Appl. Spectrosc.*, 2002. **56**: 1098–1106.
125. Taylor, R.A., *An Introduction to Statistical Quality Control*. 1987. Koch Label Co.
126. Iwamoto, R., et al., FT-NIR Spectroscopic Study of OH Groups in Ethylene-Vinyl Alcohol Copolymer. *Appl. Spectrosc.*, 2001. **55**: 864–870.
127. Miller, C.E. and D.E. Honigs, Discrimination of Different Crystalline Phases Using Near-Infrared Diffuse Reflectance Spectroscopy. *Spectroscopy (Eugene, Oreg.)*, 1989. **4**: 44–47.
128. Schmalz, H., et al., Anionic Polymerization of Ethylene Oxide in the Presence of the Phosphazene Base (Bup4)-P-T — Kinetic Investigations Using in-situ FT-NIR Spectroscopy and Maldi-Tof Ms. *Macromol. Chem. Phys.*, 2003. **204**: 1056–1071.
129. Campbell, W., The Use of Near Infrared Spectroscopy for the Determination of Hydroxyl Value in Alkoxylates, in *NIRS Int. Symposium, Technicon Industrial Systems*. 1986. Scheveningen, Holland.
130. Chalmers, J.M. and W.C. Campbell, The Use of NIRS for the Determination of Hydroxyl Value in Alkoxylates, in *Analytical Applications of Spectroscopy (Incorporating the First International Near Infrared Spectroscopy Conference)*. 1988. Norwich, UK: Royal Society of Chemistry.
131. Grady, P., *Internal Communication*. 1987, Bran and Luebbe Analyzing Technologies.
132. Walling, P., Application of NIRA in the Toiletries Industries, in *10th Int. Symposium on Near Infrared Reflectance Analysis*. 1986. Tarrytown, NY: Technicon.
133. Walling, P. and J.M. Dabney, *J. Soc. Cosmetic Chem.*, 1986. **37**: 445.
134. Pietrantonio, A., Detergents Actives and Moistures, in *4th Int. Symposium on Near Infrared Reflectance Analysis*. 1983. Tarrytown, NY: Technicon.

28 Getting NIR out of the Gutter: The Analysis of Bowling Lane Condition Using Near-Infrared Reflectance Spectroscopy

Douglas S. Burns, Alan J. Siegel, and Tom Buist

CONTENTS

28.1	Introduction	569
28.1.1	Lane Nomenclature and Surfaces	570
28.1.2	Oil Patterns	570
28.1.3	The “Tape Reader”	572
28.1.4	Op-Trac	572
28.2	Methods	573
28.2.1	Materials	573
28.2.2	Instrumentation	573
28.2.3	Measurements	574
	28.2.3.1 Calibration	574
	28.2.3.2 D Mapping	576
28.3	Results	576
28.3.1	NIR Spectra	576
28.3.2	Mapping the Entire Lane	577
28.4	Conclusions	578
	Acknowledgments	579
	Nomenclature	579
	References	579

28.1 INTRODUCTION

The following section summarizes the nomenclature and background needed to understand the lanes, oils, and measurement techniques.

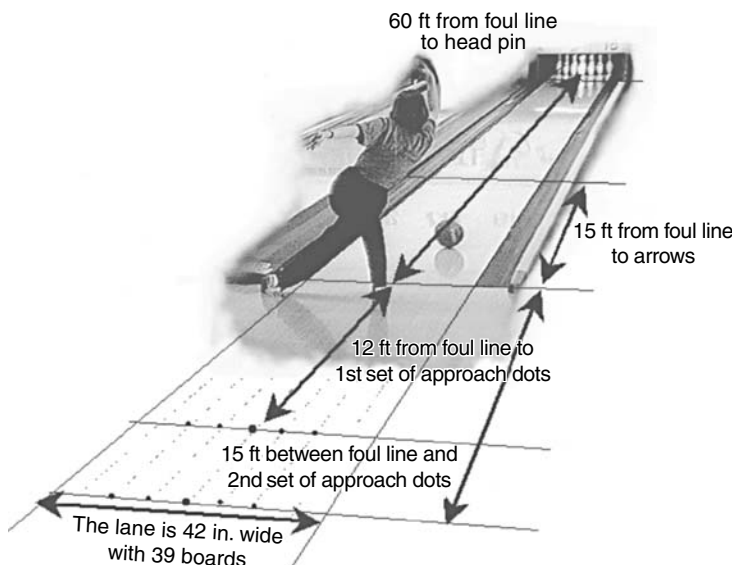


FIGURE 28.1 Nomenclature associated with the bowling lane.

28.1.1 LANE NOMENCLATURE AND SURFACES

Figure 28.1 is an illustration of the bowling lane and the nomenclature used in the sport and in this chapter. The lane is 39 “boards” wide (~42 in.) from gutter to gutter and about 60 ft long from the foul line to the head pin. Lane data are shown in terms of (a) “board number” describing the left (L2 to L19), center (C20), and right (R19 to R2) sides of the lane and (b) a distance in feet from the foul line. The “one” boards (i.e., L1 and R1) are the boards adjacent to the left and right gutters.

Bowling lanes are constructed using a variety of natural or synthetic materials. In the past, lanes were constructed using real wood boards that were inlaid and coated with various materials such as epoxy or urethane designed to protect the core materials as well as to enhance the consistency of the surface. Often a harder maple wood is used closer to the foul line and a softer pine wood is used toward the pin end of the lane.

More recently, a growing number of bowling proprietors are replacing their old wood lanes with synthetics, and most, if not all new establishments are using synthetic surfaces. One type of synthetic surface is a thin polymer laminate that serves as a protective overlay. Guardian™ is one such example and is commonly used as an affordable approach to resurfacing the lane. In contrast, various synthetic composites (e.g., ProAnvilane, GlowIQ, HPL) have been developed by various vendors to serve as replacements for old wood lanes.

28.1.2 OIL PATTERNS

A bowling ball that is thrown down the lane can exert nearly 2000 lbs of force on the surface. In order to protect the lane (as well as to provide a measure of control over the scoring condition), oil is added to the surface to reduce the friction with the ball. In the sport of bowling, the amount and distribution of oil largely controls the ball reaction to the underlying surface and largely determines whether the condition is either easy or challenging. Oil can only be applied a certain distance down the lane, and there are stringent requirements that govern the cross lane distribution of oil as well. Tournament rules govern how the oil must be applied to the lane. Two examples of oil patterns are shown in Figure 28.2, the first being a blocked pattern and the second a more gradually changing “Christmas Tree” pattern. In both cases, there is more oil applied to the middle portion of lane at the

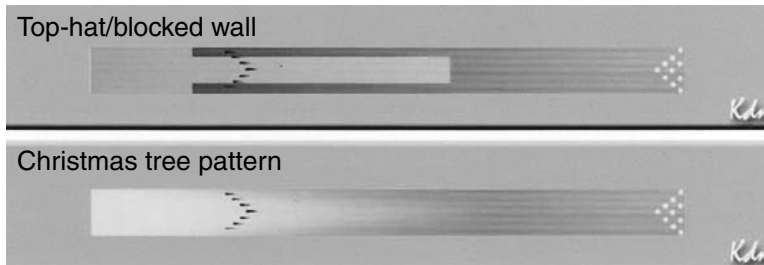


FIGURE 28.2 (See color insert following page 622.) Examples of oil patterns.

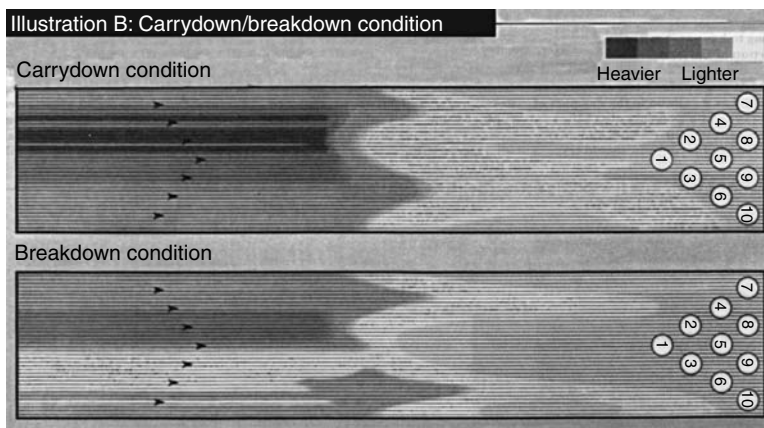


FIGURE 28.3 (See color insert following page 622.) Illustrative of carry-down and breakdown oil conditions.

foul line end, tapering gradually toward the pins. Typically, no oil is applied over the last third of the lane's surface, leaving it dry. A dry surface creates more friction between the ball and the lane, which facilitates a cross-lane migration of the ball into the pocket (the area between the 1 and 3 pins) at the back end of the alley. The length of the pattern is a major factor that controls when the ball will start to hook.

Complicating matters further, the oil applied to the surface of a bowling lane “migrates” down the lane as bowling balls travel toward the pins. The ball can push the oil around the surface or pick up the oil and drag it down the lane. This migration of the oil is known as “carry-down” and is shown in Figure 28.3. In addition, as shown here, a breakdown condition can occur in spots where the ball is repeatedly thrown. These dry spots leave the surface unprotected and further cause the lane to affect the motion of the ball. For professional or serious amateur bowlers, it is desirable to know what the initial pattern looks like and what happens to the pattern as games are played on that surface. The challenge becomes the ability of the bowler to adapt his game to the changing lane condition (i.e., moving to the left or right, expecting more or less hook on the ball).

The United States Bowling Congress (USBC), formerly known as the American Bowling Congress/Women's International Bowling Congress (ABC/WIBC), and the Professional Bowlers Association (PBA) have guidelines as to how oil may be applied to a lane surface so as to achieve a fair, but “legal” condition. Such guidelines are most strictly enforced for tournament play and “Sport Bowling” (a recent protocol that seeks to level the playing field for competitive bowling), and also are important to validating the legitimacy of perfect “300” games. These guidelines describe the amount of oil that is applied to the lane surface in terms of “units.” The current definition of a “unit” of oil varies from 0.035 to 0.116 $\mu\text{l}/\text{in.}^2$ [1]. The smaller amount deals with a film layer thickness of

~54 nm. The USBC has established a minimum oil thickness of three (3) units for a league score to be valid.

Considerable research has been devoted toward developing bowling balls, made from a variety of materials, that will react most favorably to a particular pattern of oil (or lack thereof) on a particular type of lane surface. Furthermore, ball-weighting schemes (adjusted by drilling holes to alter the weight block with respect to an internal weighting system) are also tuned to particular types of oil conditions. How to develop and select the optimum equipment for a given bowling establishment, type of lane, or generic lane condition has been the subject of millions of dollars of research and a source of considerable ongoing mystique. To make matters more complex, the pattern and quantity of oil on the lanes constantly change as bowlers, using many different tracks, redistribute the oil. Furthermore, oil is absorbed by the ball surface, causing depletion as well as redistribution of the conditioner. Unfortunately, all modern methods for analyzing the pattern of oil on a lane surface are labor intensive, time consuming, expensive and noncomprehensive. Furthermore, the entire state of the lane must be inferred and extrapolated from a finite number of discontinuous samples.

28.1.3 THE “TAPE READER”

The initial approach to facilitate “seeing” the oil on the lane surface was developed by Troxell [2]. This method required the addition of fluorescent dyes or additives to the lane conditioner. Although this technique provided an elegant qualitative picture of the concentration of additive (assumed to be equivalent to the oil distribution), it was not able to quantify the oil. The method for quantitative measurement of oil on the lane that has been used since the 1980s has been a manual oil take-up device. In order to measure the amount of oil on the bowling lane, lane oils contain “UV additives,” a fluorescent whitening agent (Uvitex OB, 2,2'-(2,5-thiophenediyl)*bis*(5-*t*-butylbenzoxazole), CAS #: 7128-64-5) used to facilitate a process called “taping” [3]. Before this research, taping was the only method of gathering any type of quantitative information about the amount of oil on the lane surface. The method involves using a device to affix adhesive tape applying constant pressure across a lane surface and lifting oil along with the UV additives onto the tape for analysis by an optical reader [4]. The method is indirect (measures additive, rather than oil), destructive (it modifies the surface by removing the oil), inherently inaccurate (the amount of oil/additive that adheres to the tape is related to the amount of pressure the lane technician applies when affixing the tape), manually tedious, relatively timely (the tape typically remains on the lane at least 10 to 15 min), expensive, and entirely impractical for routine use. Furthermore, owing to these constraints, measurements are only taken at widely spaced intervals along the length of the lane, and are, therefore, largely discontinuous. Measurements are typically taken at three distances down the lane.

Figure 28.4 illustrates representative “Tape” data obtained at three distances down the lane from the foul line for a typical “house-pattern.” Consistent with the example patterns shown in Figure 28.2, the highest oil concentrations are at 16 ft in the center of the lane. In this case, the pattern ended before 40 ft and distances from 38 ft to the pin deck had very low levels of oil. Note that the outer boards (i.e., near the gutters, L2–L6 and R6–R2) are very dry with oil levels near the required 3-unit minimum. This lack of oil in many house-patterns acts as an “artificial bumper” that helps steer the ball back toward the center of the lane. The center of the lane at 16 ft has 60 to 75 units of oil.

28.1.4 OP-TRAC

Near-infrared (NIR) was selected as an analytical approach for measuring the thickness of mineral oil-based films because it is an accepted nondestructive technique when C–H vibrations are significant. Various applications of the technology used by the Oil Pattern Topical Recognition and Control (Op-TracTM) system within the bowling industry include measurement of lane oils, lane wear, and oil adsorption into the balls and pins. Further discussion of these applications as well as the research described here have been described previously [5,6].

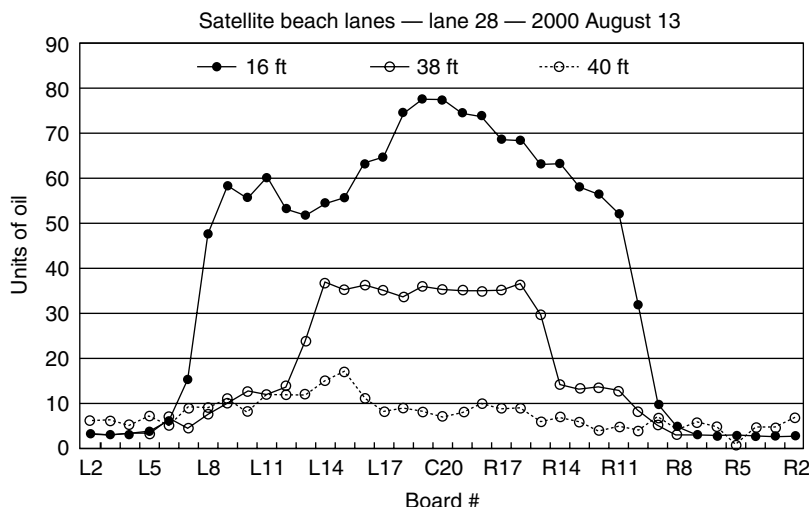


FIGURE 28.4 Representative tape readings at three distances (16, 38, and 40 ft) from the foul line.

28.2 METHODS

28.2.1 MATERIALS

The primary materials affecting the viability of an NIR-based approach to remote sensing of a bowling surface consist of a family of lane dressings (oils) and a family of lane surfaces. The types of surfaces were briefly considered in the Introduction section.

Lane oil is manufactured by several vendors. For the most part, these oils are mineral oil based with a variety of proprietary additives that determine how the oil behaves on the lane in response to substantial bowling ball traffic. Some oils are built to have higher viscosity than other oils. The carry-down characteristics of oil, the ability of the oil to flow back across ball tracks and the propensity of the oil to be absorbed by bowling balls are only a few of the factors considered in the engineering of specific oils. Some oils afford better performance on a particular type of surface, possessing properties that are conducive to high scores. Spectrally, at least in the NIR region of interest, there appears to be little difference in the treatment for the various oils. The same wavelengths seem to be relatively effective for analyzing all tested mineral oil-based lane dressings. The oil used in the work described in this paper was obtained from Kegel, Inc. (Sebring, FL) and was the Defense Oil.

The underlying lane surface appears to carry more importance vis-à-vis the portion of the NIR spectrum suitable for analysis. In general, most synthetics are relatively similar, while wood lanes with urethane or epoxy coatings have a different character. The most difficult surface appears to be Guardian, perhaps because it has significant hydrocarbon content in comparison to the other lane surfaces. It becomes a bit more challenging to differentiate oil from the laminate itself. The lane surface at the Colonial Lanes Bowling Center (Orlando, FL) and that used in the PBA tournament that is the subject of this research was a Maple with urethane coating (from 0 to 18 ft) and a Pine with urethane coating (from 18 to 60 ft). The lane was old with many imperfections from years of use.

28.2.2 INSTRUMENTATION

Diffuse reflectance measurements were made using a Foss NIRSystems Model 6500 spectrophotometer scanning from 400 to 2500 nm with a 2 nm resolution. The SmartProbe attachment was linked via a 1.5-m fiber-optic cable and mounted a fixed 1 mm above the surface. Using this configuration, light was normally directed onto the surface of the lane where it penetrated the sample

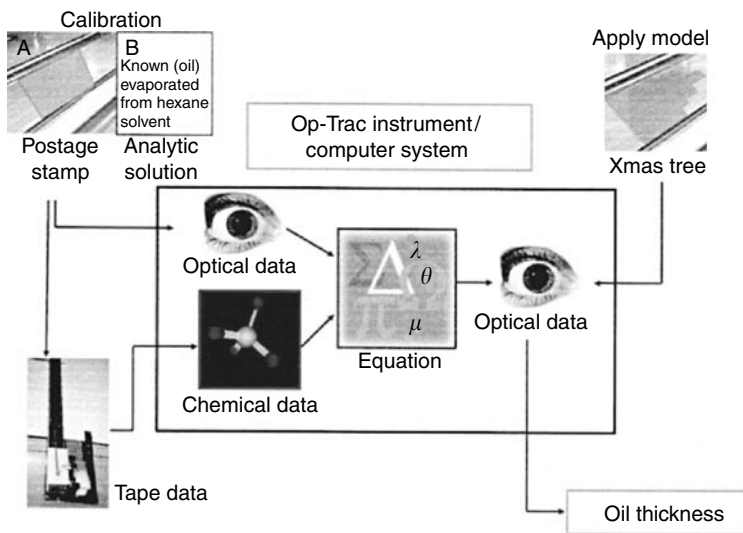


FIGURE 28.5 Overview of lane oil thickness calibration methodology.

at some distance, interacted with the sample, and was changed in some way (i.e., energy at some wavelengths was absorbed by the surface — film). The light was then reflected from the surface and measured. The “sample” in this case was the oil film and the lane surface material. Chemometric analysis of the spectra was performed using the VISION software associated with the instrument, SAS v8.1, and MS Excel.

28.2.3 MEASUREMENTS

28.2.3.1 Calibration

Two approaches were used to establish calibration data and are summarized in Figure 28.5. The initial approach (A) was to compare NIR spectral data with the taping data. The second approach (B) was to uniformly disperse a known amount of oil onto a sample lane surface. These are described in the following sections.

28.2.3.1.1 Tape Data as the Reference (A)

“Postage Stamp” patterns were placed on the lane using standard lane maintenance equipment. These postage stamp patterns are essentially rectangular patches of “known” amounts of uniformly distributed oil. In this case, the oil thickness is theoretically the same across the lane (L2, L3, . . . , L19, C20, R19, . . . , R3, R2) for some number of feet down the lane. For practical reasons there is typically a small variation in the oil thickness across the lane. Figure 28.6 presents a summary of the reference taping data obtained for three postage stamp patterns with oil dispensed at approximately 30, 60, and 80 units of oil (where 1 unit = $0.035 \mu\text{l/in.}^2$). Spectra from 400 to 2500 nm were obtained on each board and the data correlated with the oil thickness for each location within the postage stamp pattern.

Although the exact wavelengths vary slightly depending on oil/lane surface combinations, Figure 28.7 provides a general perspective as to the “important” wavelengths. These wavelengths are associated with the C—H vibrational modes of the “mineral oil” based lane conditioner. Although the total amount of oil dispensed onto the lane is accurately known, the amount in any given location is subject to significant variation (peaks, valleys, and tilt in the lane surface lead to migration of the oil and “dry” or “pooling” spots of oil). This, coupled with the error associated with the current taping

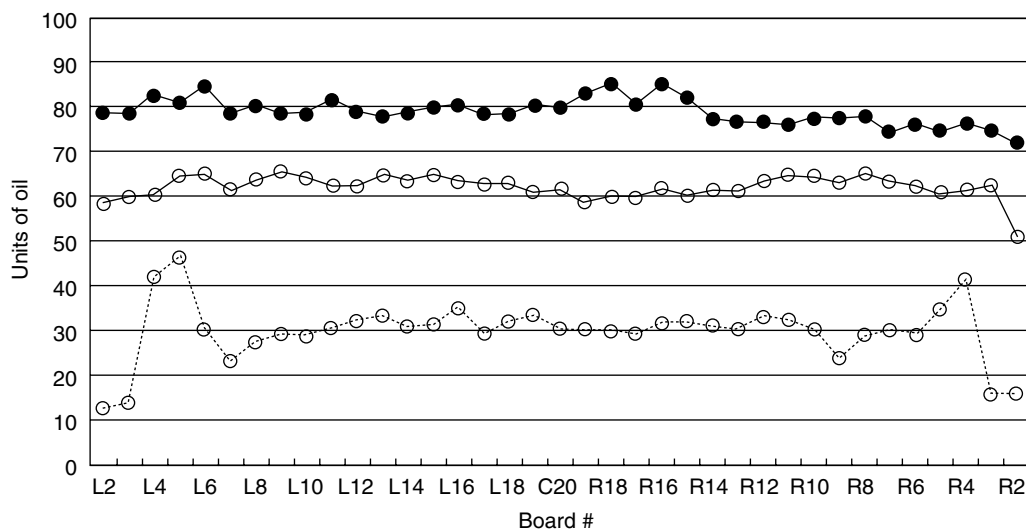


FIGURE 28.6 Representative oil thickness via taping data for three postage stamp patterns.

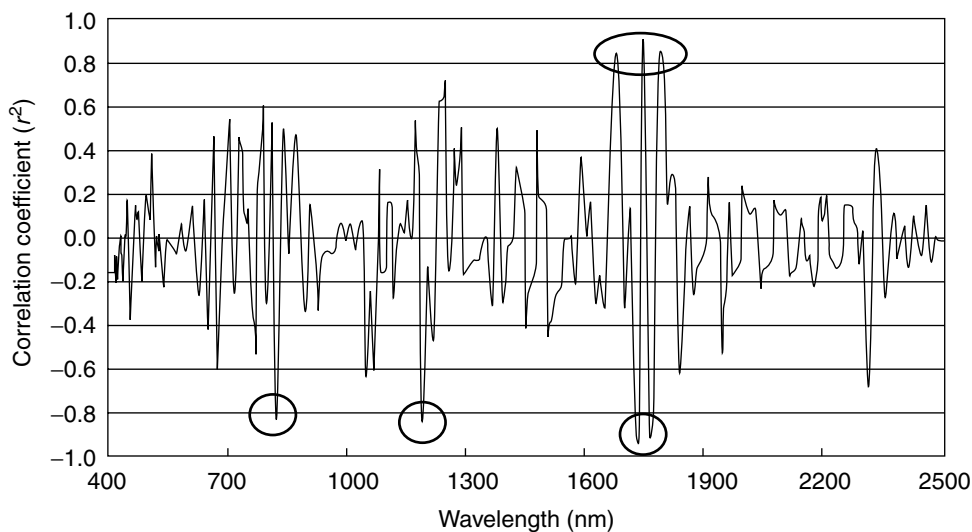


FIGURE 28.7 Correlation coefficient (r^2) as a function of wavelength for oil thickness.

method, lead to larger uncertainty in the calibrations developed from lane machines and taping data. To improve the quality of the calibration data, we therefore implemented an improved “analytical” approach.

28.2.3.1.2 Dispersing Known Amounts of Oil as the Reference (B)

To improve the quality of the calibration data, we created solutions of known concentrations of lane oil in hexane solvent. Known volumes of these analytical solutions were dispensed using an Eppendorf micropipette into samples of lane surface (1 in.²). The hexane solvent was allowed to

evaporate, leaving behind a known amount of oil that was more uniformly dispersed on a smaller sample of lane material.

28.2.3.2 D Mapping

Measurements of oil thickness were made at ~ 500 points across and down the lane. The individual points were contoured to generate a two-dimensional (2D) map of the oil thickness around the lane. Oil thickness was determined by scanning the stripped lane (zero oil) and then the oiled lane. The change in absorbance (i.e., oiled minus stripped) of the second-derivative data at 3 wavelengths (WLs) was used to calculate thickness. Even though the background (i.e., the stripped lane surface) is a significant component contained in these data, calibrations were readily obtained for measurements at a single location on the lane (i.e., the same background for a series of oil film thicknesses) without background subtraction. However, for measurements covering a large area around lane, separate background measurements were needed to be subtracted because of the significant variation in the background throughout the lane.

28.3 RESULTS

28.3.1 NIR SPECTRA

An example of the second-derivative NIR spectra from 1600 to 1800 nm as a function of increasing oil thicknesses is presented in Figure 28.8 and shows the spectral differences around 1685, 1725, 1744, 1762, and 1784 nm. Consistent with the data shown in Figure 28.7, calibration models using 3 WL were developed by choosing these and other wavelengths of high correlation (depending on the oil/surface combination).

Figure 28.9 shows an example of NIR-predicted oil concentrations using various 1 WL (open symbols) and 3 WL (filled symbols) calibration models. These data were consistent with taping data and show higher concentrations in the middle of the lane (L10 to R10) and closer to the foul line (8 ft). There is more board to board variation when using the simpler 1 WL models. The 3 WL model used here results in a smoother cross-lane profile.

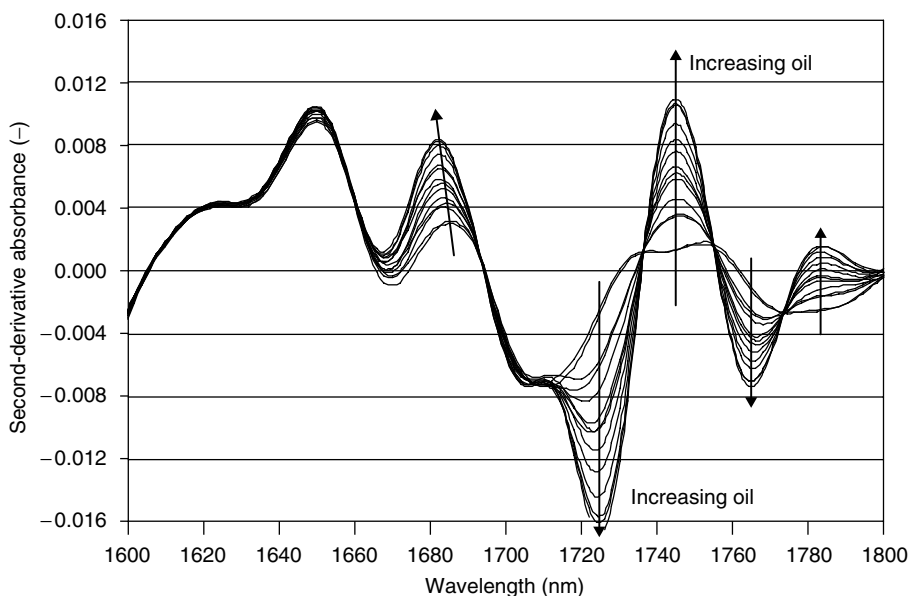


FIGURE 28.8 Second-derivative spectra of oil as a function of increasing thickness.

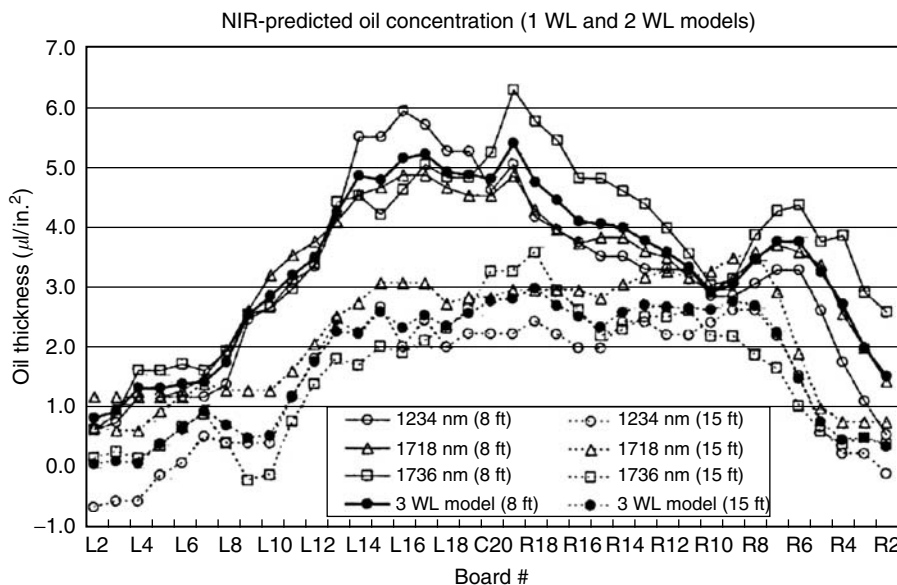


FIGURE 28.9 NIR-predicted oil concentrations.

28.3.2 MAPPING THE ENTIRE LANE

A goal of this research is to be able to provide a concentration map of the oil over the entire lane. This is accomplished by combining the various across-lane data sets (i.e., data similar to Figure 28.9) and interpolating between the points around the lane.

Figure 28.10 represents the “before” and “after” oil conditions for the PBA Regional Tour Event held during August 2002 at Colonial Lanes in Orlando, Florida. The two leftmost renderings (a and b) depict the theoretical pattern deployed by the conditioning machine, and the pre-bowling reading of the lane, respectively. Approximately 500 points were sampled to develop the data set.

The resemblance to the theoretical pattern applied to the lane by the machine is excellent. The “browns” represent the heaviest oil, followed by the “yellows.” Lighter shades of green represent minimal oil, while dark green is indicative of a dry condition. With only 500 points, it is unclear whether the indicated peaks and valleys (compared with the theoretically homogeneous pattern laid down by the conditioning machine) are real. They could be an artifact of interpolation or due to irregularities caused by “dings” in the lane where oil could preferentially puddle or run off depending upon whether the sampled location represented a hill or valley.

The “bowled” surface (part c) was sampled after the tournament’s finals, but also included traffic from the entire last set of match play (the pair of lanes used for the finals was not reconditioned after match play). Most of the right-handed bowlers began play between the fifth and tenth board to the right at the arrows. As oil was depleted and that line became less effective, the players began to play along a track that was much further to the left of their original line. By the time the final matches were being completed, players were as deep inside as the sixth arrow, about ten boards right of the left-hand gutter. The sampled pattern depicts as much, exhibiting an area of heavy wear at both the original and secondary lines for the right-handed players. One interesting feature is the oil “island” that emerges between the third and fourth arrows. This feature presumably developed as players were originally outside of this line at the beginning of match play, but then inside of this area after breakdown of the initial condition. The resulting small area of lower wear left a relative abundance of oil in this location.

The rightmost graphic shows the difference between the initial and final condition, and most explicitly displays the areas of both wear and oil “carry-down.” The greatest wear (brown area)

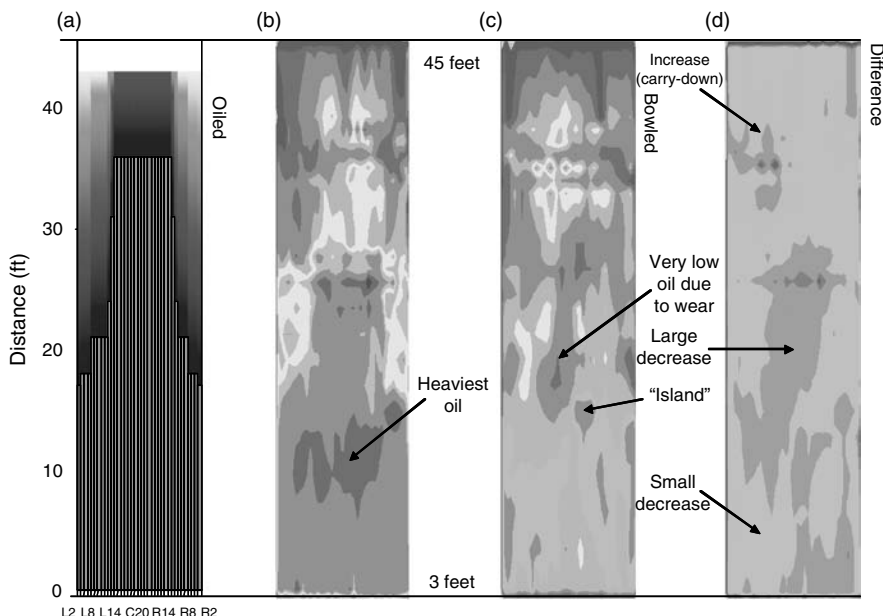


FIGURE 28.10 (See color insert following page 622.) (a) Theoretical pattern programmed into the lane oiling machine, (b) NIR-measured lane before bowling, (c) NIR-measured lane after four matches, and (d) before and after difference in NIR maps.

occurred along the deep inside line that was favored by the right-handed finalists during final match play (and also the one left-handed finalist). Maximum carry-down (green area) is noted beyond 40 ft on the left side of the lane and is probably due to a combination of carry-down from the left-hander's strike line and cross-lane spare shots by right-handers. Some carry-down is also evident along the line used to convert 6 or 10 pin spares (pins at the far right side of the pin deck). There are large areas of more general wear (yellow areas) over most of the lane. The yellow area between the green areas toward the pin end of the lane is indicative of an area of such heavy use that even the carry-down oil has been reabsorbed by bowling ball traffic.

28.4 CONCLUSIONS

This work demonstrated the potential of NIR spectroscopy to replace the existing analytical procedure used by the USBC. The NIR technology was shown to provide data equivalent to the current taping approach (cross-lane distribution of oil) for a variety of lane surfaces. The approach provides meaningful data for oil concentrations from 0 to 10 $\mu\text{l}/\text{in}^2$, which is larger than the 5 $\mu\text{l}/\text{in}^2$ maximum typically applied. Improvements to the approach (i.e., including hardware such as sources and irradiation of the surface) need to be made to improve the precision in the 0 to 0.5 $\mu\text{l}/\text{in}^2$ range. Another outcome of this effort was the realization that taping data was only of moderate use as a reference method for NIR calibration. More accurate approaches for establishing the "known" oil concentrations were required. Finally, a full lane map of the oil pattern before and after PBA tournament play was entirely consistent with the observed bowling patterns of the professionals. The 2D map enabled a quantification of the amount of "carry-down" that occurred during the tournament. Advantages of the NIR approach presented here include the following:

- *Direct measurement.* The oil is measured rather than the current indirect approach of measuring UV-active additives.
- *Nondestructive measurement.* The lane surface was measured with a noncontact approach rather than removing oil from the lane surface.

- *Results are independent of lane technician.* There is significantly reduced human error in that the measurement is independent of the lane technician. In contrast to the taping method, the NIR approach is not dependent on the pressure applied by the technician.

ACKNOWLEDGMENTS

The authors would like to thank professional bowler Mr Tom Blasco for providing lane time, contacts, and education. Work by Mr Harvey Siegel for providing engineering support to aid data collection and his significant time and effort in gathering data was most appreciated. In addition, thanks to Mr Neil Stremmel, director of research for the USBC who provided critical feedback during the research effort. Finally, the authors would like to thank the proprietors of Satellite Beach Lanes (Satellite Beach, FL), Colonial Lanes (Orlando, FL), and Kegel, Inc. (Sebring, FL) for allowing us to make measurements on their lanes.

NOMENCLATURE

2D	Two-dimensional
ABC	American Bowling Congress
mm	Millimeter
NIR	Near-infrared
nm	Nanometer
Op-Trac	Oil Pattern Topical Recognition and Control
PBA	Professional Bowlers Association
SAS	Statistical Application Software
USBC	United States Bowling Congress
UV	Ultraviolet
WIBC	Women's International Bowling Congress
WL	Wavelength
YABA	Young American Bowling Alliance

REFERENCES

1. *ABC, WIBC, YABA Complete 2004–2005 Rulebook*, 2004. Available at www.bowl.com
2. Troxell, US Patent 4,982,601, January 8, 1991.
3. Brandy Padilla, Kegel Inc., Sebring FL, Private communication, 2004.
4. "Brunswick Computer Lane Monitor™ Lane Dressing Analysis System", Application Software Version 2.0, Brunswick Bowling and Billiards Corporation, 2002.
5. Burns D.S. and Siegel, A.J. *Method and Apparatus for Monitoring Surface Condition of a Bowling Lane*, US Patent 6,794,650 B2, September 21, 2004.
6. Burns, D.S. and Schreiber, E. *Bowling Lane Condition Analysis Using NIR Spectroscopy*, Abstract 263, 40th Annual Eastern Analytical Symposium and Exposition, Atlantic City, NJ, October, 2001.

29 Process Analytical Technologies (PAT) in the Pharmaceutical Industry

Emil W. Ciurczak

CONTENTS

29.1 Introduction	581
29.2 The PAT Guidance	582
29.3 The Effect of PAT on NIR	582
29.4 Some Applications	584
Suggested Readings and References	584

29.1 INTRODUCTION

The use of near-infrared (NIR) equipment in a process setting is hardly a new concept. There is a complete chapter in this text devoted to the subject, and on-line analyses are mentioned in most of the chapters on specific industries. What is new is the application of real-time analysis and control to pharmaceutical processes. The reason for such a late introduction is due largely to the conservative nature of producing medicine.

With products such as milk (which gives off an odor when spoiled), potato chips (which are discarded if too much salt is tasted by the consumer), and an empty plastic container (if made poorly and breaks open), a consumer can protect himself from a sub-par product.

With drugs, the consumer, often ill and weak, has no warning that there is a problem with his medicine. Sub- or super-potent tablets look identical to the correct dosage level. Thus, the industry is understandably faint to attempt anything new in the production of its product. Under current testing guidelines, referred to as “current Good Manufacturing Practices” or cGMP, three production-sized batches are produced and, when shown to give satisfactory product, become the template for all future batches.

That is, the raw materials are tested for chemical purity and identity and the weights and time of blending are set (X kilos of materials, blended for Y minutes, etc.). If a wet granulation is performed, the amount of liquid added, time of mixing, and time of drying are set. If a tablet is compressed, the weight and hardness of the tablet are set. Often, only at the completion of the process are any analytical tests performed: usually only the content uniformity of the active and dissolution (release of product into solution) time.

This produces products that have been proved to be safe and effective for the public consumption for decades. But, there are a large number of problems with this approach. For one, there is a large time factor between steps, often weeks to months. It is not unusual for a single lot of product to take 4 to 6 months to produce and be approved. If there is a problem with the assay, an investigation is initiated, which in itself, may take months. In fact, audits of the industry have shown that the level of activity in production often approaches only 3%. This level would force any other industry to file

for bankruptcy under Chapter 11. The pharmaceutical industry, in part, merely charges high prices to cover the cost of manufacture. Since its products are not a luxury, consumers, often through health maintenance organizations (HMOs), pay the premium prices.

In an effort to aide this production bottleneck, the U.S. Food and Drug Agency (USFDA) issued two Guidances, designed to encourage producers to innovate and modernize how they make and analyze their product. The Guidances do not spell out *how* to use technology to understand and control the process, but are *encouraging* companies to use the best science and technologies available.

29.2 THE PAT GUIDANCE

In early 2001, Dr Ajaz Hussain (USFDA) began organizing a subcommittee to the Advisory Committee to the FDA. The purpose of the subcommittee was to solicit advice to be used to generate the PAT Guidance (final form released in September 2004). With the input from the public hearings, the FDA formulated the wording of the document. Three key passages say:

1. "PAT is a system for designing, analyzing, and controlling manufacturing processes through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality."
2. "The term *analytical* in PAT must be viewed broadly to include chemical, physical, microbiological, mathematical, and risk analysis conducted in an integrated manner. The goal of PAT is to support principles of *Quality by Design* that emphasizes fundamental process understanding and control focus to maximize process efficiency."
3. "The main focus of the Guidance is *not* the hardware, per se, but rather the concept of measurements as a means to understanding, then controlling a particular process."

It works in conjunction with the 2002 Guidance, "Pharmaceutical cGMPs for the Twenty-first Century: a Risk-based Approach." It has three goals:

1. Encourage innovation and new manufacturing technologies.
2. To focus FDA resources on areas of manufacturing considered to pose the greatest risk.
3. Improve the consistency of FDA efforts in ensuring drug quality and safety.

In other words, it is saying to the industry, "Use your best scientific judgment on a risk-based basis." While not its intention, the Guidance allows more than liquid chromatography to be used for the analysis of drug products. The effect that this Guidance and the PAT Guidance have had on all instrumentation has been remarkable. NIR has changed in a most notable manner.

29.3 THE EFFECT OF PAT ON NIR

NIR applications and hardware have been documented in detail throughout this text. The parallel evolution with other laboratory instruments has been documented. Most hardware has been developed by what could be called "traditional instrument companies." This has a dual effect: it makes continual improvement inevitable, but, because a certain number of instruments need to be sold to recoup the R&D investment, improvement is also glacial.

Most suppliers of NIR equipment are relatively small (even when larger instrument companies supply NIR hardware, their NIR division is small). This ensures that the instruments are relatively expensive, which, in turn, assures small numbers of instruments sold, which assures a continued high price.

Since, historically, NIR instruments were designed to be versatile and placed in a laboratory, the size and slow pace of change has been no problem. However, with the advent of the PAT initiative, the landscape has changed dramatically. The demand is now for smaller, faster, more specialized, often wireless instruments to be positioned along the production line.

While most other industries use process analysis equipment (and many have for decades), these industries are not as heavily regulated as pharmaceuticals. The USFDA has the job and authority to protect the American (and worldwide) consumers from faulty products. As such, they are tasked with only allowing changes that are beneficial to the process and provide necessary information. Since the overall industry (production) resembles how it appeared 50 years ago, many of the current tests are performed for “historical” and not scientific reasons.

For instance, many of the raw materials tests are “traditional” and meant to identify the chemistry and purity of the materials. Lactose, used in many formulations as a bulking agent, may be used as an example. When a sample is heated with a copper salt, a red color appears; this shows the presence of a reducing sugar. Tests such as “residue on ignition,” “heavy metals,” and “loss on drying,” while necessary to show purity of product, do nothing toward telling the formulators how well a material will work in a process setting.

Under the “PAT umbrella,” beginning with the raw materials, different tests will be administered. Using the concept of “necessary and sufficient,” tests will reveal whether a material will produce a satisfactory product. The major difficulty is that most manufacturers do not know what will produce a satisfactory product. The PAT concept is to use tests that reveal where the “critical manufacturing factors” exist. Since current tests are laboratory based, all in-line or at-line testing are “new” to the industry.

The newness of these testing points has spurred a plethora of new companies and instrumentation. In addition to traditional spectroscopic ranges, laser-induced fluorescence, thermal effusivity, and acoustics (passive and active) have made appearances. Raman has also become process-hardened, along with other techniques. In the rush to be part of the pharmaceutical process revolution, some very interesting new “players” have entered the game.

As the twentieth century came to a close, the telecommunications industry experienced a serious down-turn. The rapid growth through the late 1980s and the 1990s produced a glut in bandwidth among the large providers (MCI, Verizon, Qwest, etc.). The inevitable slowdown reached down to their technical suppliers, as well. These smaller, technology-based companies were strong on engineering and innovation in optics. To survive, they turned from communications to producing measuring equipment (spectrometers).

Using MEMS (Micro Electric Mechanical System) and SLEDs (Superluminescent Light Emitting Diodes) along with their existent expertise in fiber optics, these companies were poised to provide small, long-lived, and rugged measuring devices. When a company builds its light source to last 25 years (hard to change sources in undersea equipment), it has already made strides toward process analyzers.

These new, small companies do not have to make up costs on existing instruments, so, in essence, they manufacture “specialty” instruments for each customer using the “engines” peculiar to their company. The size and ease of installation has brought about the concept of “spare” instruments. That is, instead of waiting for a repair person from the instrument company to arrive to repair a “down” instrument, a spare or two are kept in stock. This allows a process to be delayed for a minimum amount of time, by simply swapping the spare for the problem part. This concept allows for true process control.

In addition to the hardware being introduced, there has been a continual improvement in software. Before the interest in pharmaceuticals, a tightly regulated industry, software was considered a “research tool.” That is, programs seldom had signature slots and passwords; they were “tools.” Now, with an eye to FDA approval, almost every instrument company and “third-party” software provider has provisions for security and “paper trails.” A paper trail is simply a continuous documentation of

who generated, modified, or even look[ed] at an equation, document, or spectrum. It is analogous to the “chain of custody” concept for evidence in a criminal trial.

The added attention to detail and validation of the software has given us software with far fewer problems than before the PAT initiative. Rebooting a computer at a grain elevator will not cause an “out of specification” investigation, merely inconvenience for the operator. This is yet another benefit of the guidance.

29.4 SOME APPLICATIONS

Without infringing on the other chapters, some of the applications now considered “common” are on-line blend uniformity testing. This ability to monitor a moving blender was unheard of a few years ago. The same can be said about monitoring the drying of materials in a static or fluidized bed. Not only is the solvent level measured but also any polymorphic or crystalline changes. Coating of tablets may also be followed in real time.

The power of NIR has been applied to the process line. Chemical and physical parameters are now routinely followed in real time. And, to a large degree, much of the credit goes to the FDA. Wireless instruments would probably have evolved, but not as quickly.

SUGGESTED READINGS AND REFERENCES

1. USFDA Guidance: “Part 11, Electronic Records; Electronic Signatures — Scope and Application,” August, 2003.
2. USFDA Guidance: “PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, Pharmaceutical CGMPs,” September 2004.
3. EMEA Note for Guidance on the use of Near Infrared Spectroscopy by the Pharmaceutical Industry and the Data Requirements for New Submissions and Variations, EMEA/CVMP/961/01, London, February 20, 2003.
4. USP General Chapter (1119), “Near-Infrared Spectrophotometry,” Official 8/1/04.
5. USFDA Guidance: “Pharmaceutical cGMPs for the Twenty-first Century: a Risk-based Approach,” 2002.

30 Pharmaceutical Applications of Near-Infrared Spectroscopy

Carl A. Anderson, James K. Drennen, and Emil W. Ciurczak

CONTENTS

30.1	Introduction	585
30.2	Applications in the Analysis of Tablets and Solid Pharmaceutical Dosage Forms	586
30.3	Determination of Tablet Actives by Near-Infrared Spectroscopy	587
30.4	Foundations of Analysis of Solid Dosage Formulations	588
30.5	Analysis of Solid Dosage Formulations during Manufacturing	594
30.6	Analysis of Intact Dosage Forms	596
30.7	Hardness	601
30.8	Considerations for Intact Dosage Form Analysis	603
30.9	Near-Infrared Imaging	604
30.10	Conclusions	605
	References	606

30.1 INTRODUCTION

Although the pharmaceutical industry was slow in adopting this technique, a number of pharmaceutical applications of near-infrared (NIR) have been reported. This chapter will discuss the development, application, and validation of NIR spectroscopy for the pharmaceutical industry; the software and multivariate methods used for this work are covered extensively in other chapters and will be mentioned in passing.

Though reported by Herschel in 1800, this spectral region was largely ignored until the late 1950s. The first publications describing pharmaceutical applications of NIR spectroscopy appeared approximately 10 years later. Several articles on this topic were published during the 1970s and early 1980s; the late 1980s brought a distinct increase in the frequency of published articles that, again, increased sharply in 2006 (see Figure 30.1). This reflects the growing popularity and increasing maturity of the technique for pharmaceutical analyses over the past few decades with the marked increase of the mid-2000s resulting from the process of analytical technology initiative from FDA [1]. General reviews of NIR spectroscopy have been published [2–8] and contain references to a number of earlier reviews of the technique and applications.

Several texts on NIR analysis are also available [9–16]. Three of the texts, those edited by Patonay, by Burns and Ciurczak, and by Ciurczak and Drennen contain chapters dedicated to pharmaceutical applications of NIR spectroscopy. Ciurczak also authored a comprehensive review of

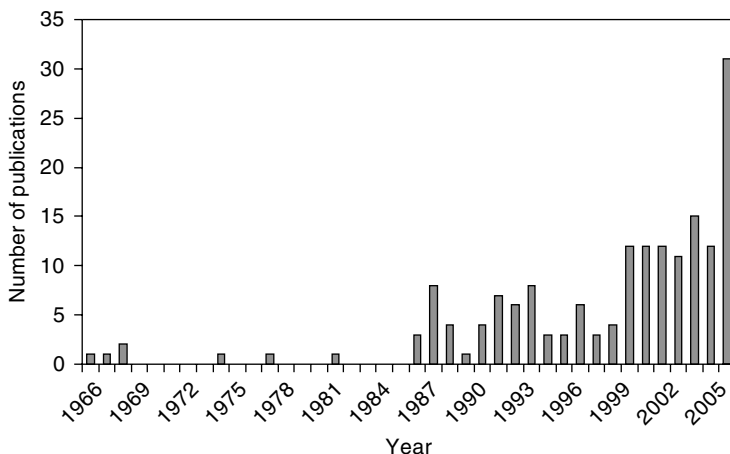


FIGURE 30.1 Number of pharmaceutical publications utilizing NIR spectroscopy.

the pharmaceutical applications of this method, and other papers discussing selected topics in NIR analysis of pharmaceuticals have been published [2,17–20].

The NIR region of the electromagnetic spectrum is generally defined as between 700 and 2500 nm. Absorption bands in this region are due to overtones and combinations of the fundamental mid-IR bands. Most compounds have low molar absorptivities in the NIR region, exhibiting broad, overlapping absorbance peaks. The low absorptivities, once considered a shortcoming of the NIR region, have become a primary reason the method is used extensively in the pharmaceutical industry. NIR absorbances arise primarily from anharmonic oscillations arising from C–H, O–H, and N–H bonds, common to most drugs.

Several qualities of the NIR method have made it appealing as an alternative to traditional analytical techniques for pharmaceutical products: samples can be scanned as is, requiring little or no preparation prior to analysis; analyses of complex matrices are performed rapidly, with results often obtained in a few seconds or less; and, unlike many other methods, NIR spectroscopy does not use expensive, dangerous solvents. These attributes make the technique well-suited for routine laboratory and process control applications.

30.2 APPLICATIONS IN THE ANALYSIS OF TABLETS AND SOLID PHARMACEUTICAL DOSAGE FORMS

The earliest publications of NIR assays of pharmaceuticals by NIR appeared in the late 1960s. In most cases, the drug was extracted from the dosage form and analyzed in solution. In several of these early studies, spectra of solid state samples of drug reported.

In 1966, Sinsheimer and Keuhnelian [21] investigated a number of pharmacologically active amine salts both in solution and in the solid state by NIR spectroscopy. Quantitative analysis of the samples in solution was performed using the 2150 to 2320 nm region. Solid-state samples consisted of compressed pellets composed of an amine salt/KCl mixture and pellets of amine salt only. Spectra of the solid samples collected in the 1050 to 2800 nm region were analyzed qualitatively using peak assignments. Several spectral features were noted as showing promise for the quantitation of drugs in the solid state, although no calibrations were developed. This was the first report of the analysis of pharmaceutical solids by NIR.

In 1967, Oi and Inaba [22] published an article on the quantitation of allylisopropylacetureide and phenacetin in pharmaceutical preparations. The samples were dissolved in chloroform and the concentrations determined using absorbance values at 1983 nm for allylisopropylacetureide and

2019 nm for phenacetin. Concentrations of these two drugs in standard mixed samples and unknown samples were determined without interference from three other drugs present in the formulation.

In another application, Sinsheimer and Poswalk [23] investigated the technique for the determination of water in several matrices. Solid-state samples were analyzed, in which hydrous and anhydrous forms of strychnine sulfate, sodium tartrate, and ammonium oxalate were mixed with KCl and compressed into disks containing 100 mg KCl and 25 mg of drug. In some samples, the water absorbance band at 1940 nm was clearly seen in the spectra of the hydrates. Quantitation of water in these samples was not successful. In other samples, differences in the spectra above 2100 nm were not explained.

30.3 DETERMINATION OF TABLET ACTIVES BY NEAR-INFRARED SPECTROSCOPY

In the earliest NIR assays of solid dosage forms, tablets and capsules served as the starting point, but were not analyzed intact. The drugs were extracted from the dosage form and the concentrations determined in solution.

The first use of NIR for tablet drug content was reported in 1968 by Sherken [24]. In this study, an assay for meprobamate in tablet mixtures and commercially available preparations was established. A range of standard solutions of meprobamate analyzed by the NIR method was used for calibration development. He used the two wavelengths corresponding to the symmetric and asymmetric stretching modes of the primary amine group of the drug molecule. The new method was as accurate, but less tedious than the official assay and recommended as the official method for determination of meprobamate.

In a study conducted by Allen [25], NIR was used for the determination of carisoprodol, phenacetin, and caffeine content in tablets. Twenty tablets were pulverized and an aliquot, dissolved in chloroform. A series of standard solutions of carisoprodol, phenacetin, and caffeine were scanned between 2750 and 3000 nm. All exhibited Beer's law behavior (maximum 200 mg/ml). Carisoprodol and phenacetin were determined simultaneously at 2820 nm for carisoprodol and 2910 nm for phenacetin. Caffeine concentrations were determined at 3390 nm. Since chloroform is a strong absorber, carbon tetrachloride was substituted as the solvent. The NIR method provided a coefficient of variation (CV) for all three drugs of 1.4% or less.

In many of these early examples of pharmaceutical analysis by NIR, the investigators worked in the 2800 to 3100 nm region, considered to be part of the mid-IR region. However, these were pioneering efforts in NIR pharmaceutical analysis. The NIR analyses of pharmaceuticals reported since these early studies have generally been performed in the 1100 to 2500 nm region, primarily owing to the design of currently marketed instrumentation.

In 1977, Zappala and Post [26] investigated the use of NIR in the analysis of meprobamate in four pharmaceutical preparations: tablets, sustained-release capsules, suspensions, and injectables. By the publication of the paper, a colorimetric method for the assay of meprobamate in tablets had been adopted in USP XIX. This colorimetric method was more rapid than the previous assay, but still required close control of reagent pH. NIR remained an attractive alternative for determination of meprobamate concentrations in dosage forms.

This new NIR method was an improvement over the NIR method introduced by Sherken nearly 10 years earlier. It took advantage of a meprobamate primary amine combination band at 1958 nm, not subject to the interference that the peak at 2915 nm suffered. The previous NIR method required removal of an alcohol preservative from the chloroform prior to analysis that interfered with the 2915 nm drug absorbance peak. The 1958 nm absorbance band, though weaker, was sufficient for calibration development.

The new method required 20 tablets or capsules to be pulverized and an aliquot of meprobamate dissolved in chloroform. Testing the calibration, nine commercial products from four different

manufacturers were analyzed. The CV was 0.7% for tablets and 1.3% for capsules, compared to 1.5% for the reference method.

Used as described above, NIR has a detection limit of 0.1% for solids. High-potency drugs constitute a small percentage of the dosage form, making accurate quantitation of drug by NIR difficult. In a 1990 paper by Corti et al. [27], an extraction procedure was used to improve the detection limit of oral contraceptives, small tablets with a low-concentration of drug.

Ethinylestradiol and norethisterone, two synthetic hormones commonly used in oral contraceptives were analyzed in this study, both qualitatively and quantitatively. Tablets with a total mass of 80 mg, 0.05 mg ethinylestradiol content, and 0.25 mg norethisterone content were used. Drug was extracted from the powdered samples with chloroform. Aliquots of the extracts were placed in an aluminum sample cup and scanned in triplicate with a Technicon InfraAlyzer 450 spectrometer.

For qualitative analysis, six wavelengths were used in the Mahalanobis distance calculation. The program was able to discriminate between samples, and to distinguish the ethinylestradiol extracts even at concentrations below 0.05%. For the quantitative analysis, a method based upon multiple linear regression was employed. Twenty extracts were used for calibration development, with ethinylestradiol and norethisterone concentrations varying over a 10% range. The correlations obtained for the two calibrations were high (ethinylestradiol calibration $r^2 = .85$ and norethisterone calibration $r^2 = .86$). Given the low-drug concentrations in the samples, however, the standard errors of calibration were also high.

The researchers attempted to use the calibrations for the prediction of samples not extracted from the tablets with a calibration based upon tablets prepared in the laboratory. When the tablet-based calibrations were used to determine drug concentration in the synthetic samples, poor results were obtained. This was an early illustration of the importance of the similarity between calibration and test samples when an accurate prediction is required.

30.4 FOUNDATIONS OF ANALYSIS OF SOLID DOSAGE FORMULATIONS

The next development in NIR spectroscopic analysis of pharmaceuticals was the analysis of solid dosage formulations. This represented a significant advance in pharmaceutical analysis, because it increased the potential for the application of the method to a large number of pharmaceutical processes. NIR was no longer restricted to the analytical laboratory; rather, it could now be used on the production floor. In addition, it obviated the need for extraction with solvents like chloroform and carbon tetrachloride prior to NIR analysis. This technique has been used in the NIR analysis of blends, granules, encapsulation matrices, and milled tablets.

An early paper using NIR in the analysis of pharmaceutical mixed powders was published in 1981 by Becconsall et al. [28]. NIR and UV photoacoustic spectroscopies were employed in the analysis of propranolol/magnesium carbonate mixtures. In this study, complete spectra were collected in the 1300 to 2600 nm region, using carbon black as the reference sample. An aromatic C—H combination band at 2200 nm and an overtone band at 1720 nm were used in the quantitation of propranolol in the drug-excipient mixtures.

The composition of the mixtures was varied over a wide range and a spectrum collected at each concentration. While the calibration developed with UV exhibited nonlinearity, the calibrations obtained using the two NIR wavelengths were linear. This linearity in the NIR region was attributed to the decreased light scattering efficiency at longer wavelengths and the more equal absorption coefficients of propranolol and magnesium carbonate in this region.

The authors concluded that pharmaceutical quality control measurements could potentially be applied in the NIR region, but cautioned that spectral interferences from other components in the sample matrix could complicate attempts at quantitation.

From 1982 through 1985, little was published on NIR analysis of solid dosage forms. Since 1986, the rapid growth in popularity of NIR for the analysis of solid dosage forms has been reflected by the numerous publications on this topic.

The first was a 1986 paper by Ciurczak and Maldacker [29]. These investigators used NIR in the analysis of tablet formulation blends and examined three methods of data treatment: spectral subtraction, spectral reconstruction, and discriminant analysis. Blends were prepared in which active ingredients (aspirin, butalbital, and caffeine) were either omitted from the formulation or varied over a concentration range of 90 to 110% of labeled strength. All samples were ground in a powder mill to ensure homogeneity, then scanned in a sample cup on a model 500C InfraAlyzer.

The first experiment examined the use of spectral subtraction. Spectra of true placebos were subtracted from spectra of the complete formulation, yielding spectra very close to that of the omitted drug. This technique could be used in the identification of constituents present in a tablet matrix.

The second experiment also involved the qualitative identification of formulation constituents. Spectral reconstruction was performed with commercially available software. Using a series of mixtures of known concentrations, the program used modified multiterm linear regression equations to correlate spectral changes with drug concentration changes. The spectrum of the drug could then be reconstructed, providing identification of drugs present in the sample blend.

The third experiment involved the classification of samples by discriminant analysis. In one series of blends, the caffeine, butalbital, and aspirin concentrations were varied independently between 90 and 110% of labeled strength. In another series, one of the three drugs was excluded from the mixture, while the other two were varied between 90 and 110%. The Mahalanobis distance calculation was employed successfully in the classification of formulations into similar groups. This technique was also used in the analysis of samples from complete formulations (in which all three drugs were present at 100% of labeled strength), borderline formulations, and samples which lacked one of the three active components. Although all the samples were classified correctly by the software, the authors noted that the spectra were nearly indistinguishable, forcing the analyst to depend on the software, rather than visual spectral differences, to make the identification.

Chemometric techniques are excellent for differentiation between samples with subtle spectral differences. However, relying on these routines without an understanding of the underlying physical and chemical phenomena responsible for the spectral differences could lead to unstable calibration models or classification based upon extraneous spectral characteristics.

These experiments highlight another advantage of NIR in dosage form analysis. For products containing multiple active ingredients, individual reference tests must generally be conducted to verify that each ingredient is present in its correct concentration. NIR can be used to characterize the entire formulation simultaneously, making multiple reference tests unnecessary.

The following year, Ciurczak and Torlini [30] published a paper on the analysis of solid and liquid dosage forms by NIR. The authors contrasted the use of NIR in the development of calibrations for natural products with that for pharmaceutical dosage forms. They noted that natural products require the development of a reference method for calibration development, while pharmaceutical dosage forms could be synthetically prepared to develop a calibration.

One disadvantage to this technique is the fact that samples prepared in the laboratory are often spectrally different from production samples due to differences in preparation methods, processing equipment, or batch size, making calibration by this method more difficult. In other words, this approach to calibration development often leads to unsatisfactory results, and the use of production samples for calibration development is preferred when production samples will be tested. In either case, a calibration range wider than that expected for production samples is necessary, to minimize the likelihood of accepting a false sample. This subject will be discussed in greater detail later.

In the study by Ciurczak and Torlini, the performance of NIR was compared with high performance liquid chromatography (HPLC) for speed and accuracy of results in dosage form analysis. The effect of milling the samples prior to analysis was also investigated. Two solid dosage form matrices,

a caffeine–acetaminophen tablet mixture and an acetaminophen tablet mixture, were prepared. Acetaminophen tablet mixtures were analyzed after milling, and caffeine–acetaminophen tablet mixtures were analyzed with and without milling. The samples were placed into a sample cup and scanned by a model 500C InfraAlyzer. Multiple Linear Regression was used for the calibration with 80% of the samples used for calibration development and the remainder for validation.

The investigation of the effects of milling showed that this sample preparation procedure did not necessarily improve results. Milling of the caffeine–acetaminophen mixture significantly improved the determination of acetaminophen, but the determination of caffeine was virtually unchanged. The largest differences between the NIR determined concentrations and theoretical concentrations were attributed to sample handling variability.

The average difference between the theoretical and predicted drug concentrations was approximately 0.25%, competitive with HPLC determinations. The NIR method also provided the advantages of rapid analysis times and no solvent purchase or disposal costs.

Granules can be used as a final dosage form, or as a processing intermediate requiring tableting or encapsulation. In a 1987 paper, Chasseur used NIR for the assay of cimetidine granules [31]. Batches of granules were prepared in which the cimetidine concentration ranged from 70 to 130% of label strength; spectra were collected on a Pacific Scientific filter-based instrument. For calibration development, both first and second derivative spectra were analyzed, with one or two wavelengths included in the model. A two-wavelength model using first derivative spectra provided the best results, giving a standard error of 1.75%.

These researchers also compared the performance of NIR with the UV reference method in the determination of cimetidine content. Samples of 100% label strength were prepared and analyzed by both methods. The standard errors for the NIR (2.73%) and the UV (2.97%) assays were virtually identical.

In a 1987 paper, Osborne investigated the use of NIR spectroscopy in the determination of nicotinamide in vitamin pre-mixes [32]. Although this was not specifically a tablet matrix, the study showed potential advantages of the method for pharmaceutical analysis. HPLC was the existing reference method for nicotinamide, requiring 3 days to analyze 36 samples. The proposed NIR method required only 30 min to analyze the same samples.

Twenty-five mixtures were used for calibration development, with nicotinamide concentrations ranging from 0 to 6% of the pre-mix formulation. Spectra were collected in duplicate on a Pacific Scientific Mk I 6350 at 2 nm intervals between 1200 and 2400 nm. The spectra were reference corrected using a standard ceramic reference. Second derivative spectra were calculated and the calibration obtained by regressing nicotinamide concentration against the ratio of the second derivative values at 2138 nm (a nicotinamide absorbance peak) and 2070 nm (a spectral minimum in the pre-mixes). The validity of the calibration was determined in two ways: first, 25 more samples were prepared in the laboratory and used to validate the model; second, 25 commercially prepared pre-mix samples with different batches of raw materials were analyzed by the NIR method and by the HPLC reference method.

The standard error of prediction (SEP) for the validation set was 0.56% w/w, with sample positioning variability accounting for half of the total sample variance. To further validate the calibration, commercial samples were collected over a period of several months and analyzed by HPLC and NIR. The results obtained by the two methods were not significantly different, verifying the NIR calibration.

In NIR calibration, the introduction of unexpected components (e.g., incorrect chemicals or contaminated raw materials) into a sample may go undetected, causing erroneous results. This type of false sample is potentially more serious than samples in which the correct constituents are present in the wrong concentrations. In a 1988 paper, Lodder and Hieftje [33] discussed in detail the application of a new algorithm, the quantile-BEAST (Bootstrap Error-Adjusted Single-sample Technique), which was more sensitive to false sample detection. The quantile-BEAST was proposed as a new method to assess pharmaceutical powder blends qualitatively.

The quantile-BEAST algorithm is a nonparametric bootstrap method based upon the work of Efron. In the 1988 study, four individual benzoic acid derivatives were analyzed as were mixtures of the four derivatives. The active ingredient concentrations were varied between 0 and 25% of the sample, with aluminum oxide used as a diluent. The samples were ground and passed through a 100-mesh screen prior to analysis. Spectra were collected in triplicate at three wavelengths on an InfraAlyzer 400.

After analysis by the bootstrap algorithm, the individual benzoic acid derivatives were classified into clusters using the measurement of nonparametric standard deviations (SDs), analogous to SDs in parametric statistics. A value of three SDs or less was expected for identical samples, while dissimilar groups were expected to have SDs greater than three. The four derivatives displayed a minimum distance of 39 SDs from each other. A “worst-case” analysis was conducted, in which acetylsalicylic acid was mixed in with the formulations at concentrations between 1 and 20%. Although three of the ten contaminated samples failed the three SD test for being false, none of the uncontaminated samples were incorrectly identified. All the compounds, including the contaminant, were closely related derivatives of benzoic acid, making their differentiation and identification at relatively low concentrations even more notable.

The most important factor in the accuracy (bias) of this method was the size of the training set, while the most important factor in determining the CV was the number of bootstrap replications of the training set.

Polymorphism is a phenomenon of great concern to pharmaceutical formulators. A number of drugs have been shown to undergo polymorphic transformation under compression, creating potential stability and bioavailability problems. Gimet and Luong [34] published results of a study in which NIR spectroscopy was used for the quantitative determination of polymorphs in a formulation matrix. Most polymorphs exhibit spectral differences in the mid-IR, and, since NIR spectra arise from overtones and combinations of mid-IR absorbances, NIR was also reasoned to be suitable for the analysis of polymorphs. An investigational drug was shown to have two polymorphic forms, with the more stable form transformed to a less stable polymorph under pressure. The researchers employed NIR spectroscopy during formulation development to quantify the more stable polymorph.

Spectra of the polymorphs displayed minor variations, but no distinguishing characteristics were obvious. The authors did not attempt to identify spectral differences according to polymorphic changes in crystalline structure. Had difference spectra (in which the spectra of one polymorphic form are subtracted from the spectra of the other form) been calculated and studied, the spectral variations between polymorphic forms may have been enhanced, providing further insight into the phenomena which allow polymorphic forms to be discernable by NIR spectroscopy. Further research should be conducted into this application of NIR spectroscopy.

Quantitative determination of indomethacin was studied by Otsuka et al. in 2003. Tablets containing α [alpha] and γ [gamma] forms of indomethacin were manufactured at laboratory scale. The sum of the two polymorphs in the tablets constituted 50% of the tablet. A calibration was developed to predict the percentage of γ of indomethacin in the tablets from 0 to 100%. Powder x-ray diffraction was used as the reference method for the determination of polymorph content.

Changes in the NIR spectrum that are indicative of the presence of a polymorphic form are frequently subtle, but are detectable with the proper processing of the spectra. In 2005, Li et al. [35] demonstrated the detection of a second polymorph (form B) of an API (proprietary) in a formulation following wet granulation. Subsequent spiking studies with the discovered polymorph over a range of 1 to 8% enabled researchers to quantitatively predict the fraction of the API that had been converted to form B during wet granulation.

Qualitative determination of the polymorphic form of an API is achievable by NIR. In 2006, Blanco et al. [36] demonstrated that the polymorph of the API used to manufacture a tablet was detectable in the finished dosage form. Thus, demonstrating the ability of NIR spectra to verify the correct form of the API was utilized in the manufacturing process. NIR was used to determine which crystalline form (A or B) of formulations of desketoprofen trometamol (DKP, 10 to 25% w/w

of dosage form) was used to manufacture a tablet. Investigations indicated that different manufacturing processes had the potential to modify the crystalline form of DKP which were detected by NIR.

The ability of NIR to detect changes in the API can facilitate investigations into quality failures. Bauer et al. [37] utilized NIR to detect the desolvation of erythromycin in a finished tablet following dissolution failures. NIR was used to follow the loss of water from erythromycin and the uptake of those waters of hydration by $\text{Mg}(\text{OH})_2$. Although the method describes on the ability to monitor the reaction in a qualitative sense, the potential for quantitative determination of the exchange of water, and the subsequent prediction of dissolution behavior is apparent. This work also demonstrates the ability to follow a solid-state reaction by NIR without specific knowledge of the specific reaction species.

The application of NIR analysis to the quantitation of ketoprofen in a gel matrix and a powder matrix intended for encapsulation was reported by Corti et al. in 1989 [38]. The goal of the research was to compare the utility of a calibration developed over a wide concentration range with that of a calibration developed over a narrow range in the prediction of unknown samples.

The NIR spectrometer was a filter-based Technicon InfraAlyzer 450 and a sample cup was used to hold the powdered samples. Reproducibility of the assay was checked by scanning a single sample in one sample cup ten times; the CV was 0.941%. When the same sample was scanned in ten different sample cups, a CV of 1.214% was obtained.

For the narrow calibration range, 13 samples were prepared that fell within $\pm 5\%$ of the theoretical content (33% active) of the matrix. The wider calibration range consisted of the original thirteen samples and an additional seven samples whose ketoprofen concentration ranged from 3 to 30%.

Several characteristics of the two calibration approaches were noted. First, the SEP for both calibrations was approximately 2%, with no sample having an error greater than 3.5%. Second, the wider calibration did not offer any significant advantages, as neither the standard error of estimate (SEE) nor the SEP improved. When the samples to be analyzed fell into a narrow range of concentrations (typical of production samples), relatively few samples were needed to develop a calibration. When the potential for considerable variability of the matrix exists (such as in the NIR analysis of natural products), many more samples are needed to generate a valid calibration.

In the analysis of powders in which the ingredient of interest varies over a wide concentration range, deviations from linearity are possible and can be a significant source of calibration error. Researchers should consider this phenomenon when developing such calibrations. For example, even though it might be valuable to use a calibration ranging from 50 to 150% of desired active ingredient concentration for quantitation of drug in both good and bad samples, it is unlikely that a linear response would be obtained over such a large concentration range. The range over which linearity is possible will depend on the concentration of drug in the matrix and the nature of the dosage form matrix itself. Additionally, the range of a calibration may significantly affect the commonly reported chemometric statistics such as root mean square error of prediction (RMSEP). For calibrations of a wide range, the RMSEP may be acceptably low, but the ability to distinguish between similar concentrations may be absent.

The following year, Corti et al. [39] applied NIR reflection spectroscopy to the analysis of ranitidine and water content in tablets. These were production samples with a narrow distribution of concentrations, so the authors established two criteria for the calibration development. First, the narrow range of sample concentrations allowed fewer samples to be included in the calibration set. Second, production samples had a normal concentration range of 4%, and the calibration was expanded to cover a 10% concentration range. These samples were prepared by adding either filler or drug to a commercial mixture (a spiking study).

Near-infrared spectra were collected on powdered samples, obtained by crushing five tablets and underdosing or overdosing the samples as described. A Technicon InfraAlyzer 450 filter-based instrument was used for the analysis. Tablet drug content was determined by HPLC and water content by Karl Fischer titration.

For quantitation of tablet drug content, three calibrations using multiple linear regression were developed. The first used only laboratory samples and provided a low SEE, but a SEP of 8.4% when the unknown commercial samples were tested. The second calibration used only production samples and provided a SEP of 1% for production samples; the SEP was 6.4% for laboratory samples. The third calibration, built with production and laboratory-modified samples, had SEPs of approximately 1% for both production and laboratory samples. The results showed that progressive widening of the calibration range provided an advantage only up to a point, beyond which no improvement in SEE or SEP was seen. This optimum value was determined to be a calibration range of approximately 5%.

The calibration for determination of water content also employed production samples and modified production samples. Both SEE and SEP were less than 0.1%. In a test of production samples over 1 year, the NIR had the greatest error in moisture prediction of tablets with less than 1% moisture. When used qualitatively, in no case did the NIR method erroneously reject samples with a moisture content greater than 2%, the moisture level at which tablets were rejected. As in the previous publication, the results supported the hypothesis that for products with little variability, a small number of samples (~10 to 20) is sufficient for calibration development.

One study in which NIR was investigated and not found to be a suitable analytical method was reported by Ryan et al. in 1991 [40]. The purpose of the study was to find a rapid and routine method for the verification of the correctness of clinical packaging. Both mid-IR and NIR were investigated for the analysis of two structurally similar cholesterol-lowering drugs, lovastatin, and simvastatin. For this study, a Pacific Scientific Model 6250 and a Bio-Rad FTS 40N were employed, with KBr powder used as a reference. Sample preparation involved grinding the samples to a uniform particle size prior to analysis. Both mid-IR and NIR offered a detection limit of approximately 1% (w/w) in the presence of excipients. NIR was unable to differentiate between the two drugs at low concentrations. Since it was conceivable that these two compounds would be present in low concentrations in the same clinical packaging, the NIR technique was discontinued. However, as reported in another chapter, this type of detection has been accomplished [41–43].

In 1992, Corti et al. [44] analyzed a variety of antibiotic compounds by NIR, including primary materials, partially processed granules, and an antibiotic cream. Spectral collection was performed on an InfraAlyzer 450. All samples were read in triplicate, with the sample cup emptied and refilled between readings. As in the authors' earlier work, multiple linear regression was used for the calibration development and Mahalanobis distances for qualitative analysis.

Qualitative analysis involved the differentiation among ten antibiotic preparations, including three types of ampicillin (amorphous, crystalline, and trihydrate) and blends of erythromycin powder and granules. All samples tested were easily differentiated based upon Mahalanobis distances. Calibrations were developed for each material, including 15 samples spanning a 5% concentration range. The SEE and SEP for each calibration were less than 2% and most were less than 1%.

A 1993 paper by Blanco et al. [45] addressed some of the concerns regarding the laboratory manipulation of production samples prior to NIR analysis. In this study, two different commercial preparations of ascorbic acid (vitamin C) were analyzed, one a granular product, the other an effervescent tablet. No less than five batches were used in the calibration development for each product.

Samples for calibration and validation were ground to a specific mesh size (250 or 100 μm) before analysis. To expand the calibration range, samples were either diluted, with the primary inactive ingredient added to each, or overdosed with ascorbic acid. Using calibration matrices containing increasing numbers of overdosed samples and comparing the sum of squared differences between the laboratory-determined and NIR-determined values, the authors verified that the underdosing and overdosing process did not affect the physical properties of the samples. Avoiding calibration problems in dilution or concentration of samples is possible by milling samples to a uniform particle size.

Spectra of the homogenized samples were collected in triplicate on a NIRSystems 6500 in reflectance mode. The effects of sample particle size on NIR spectra are often significant, but can

usually be minimized by preprocessing the spectra. Prior to analysis, three preprocessing methods were evaluated: multiplicative scatter correction (MSC), signal scaling, and first derivative. In this study, the first derivative spectra provided the best calibration results and were used throughout the experiment.

Two chemometric techniques were used for calibration development. The first was stepwise multiple linear regression (SMLR), using four wavelengths or fewer, and provided correlation coefficients greater than or equal to .99. The SEE and SEP for all calibrations were less than 2.4%. Calibrations were also developed using partial least squares (PLS) regression, on both full (1100 to 2500 nm) and reduced wavelength (1300 to 1800 nm) spectra. Generally, two or three factors were adequate for the calibrations and most of the models had SEE and SEP values less than 2%. SMLR gave more accurate results for the simpler granule preparation, while PLS was more accurate in drug quantitation for the more complex effervescent tablet formulation.

These reports of NIR analysis of blends and granules lend substantial support to the usefulness of this technology in the monitoring of intermediate pharmaceutical processes, such as blending, granulation, and qualification of bulk powders prior to tableting or encapsulation. NIR analysis of powdered tablets and capsules takes advantage of the analytical speed of this method, but does not capitalize on the nondestructive nature of the technique or the versatility of its sampling capabilities. Over the past 20 years, research has first intensified in the area of intact dosage form analysis and subsequently attention has been turned to use of NIR as a tool for monitoring pharmaceutical manufacturing.

30.5 ANALYSIS OF SOLID DOSAGE FORMULATIONS DURING MANUFACTURING

The monitoring of pharmaceutical manufacturing process, as they occur has become a primary goal of many analytical science practitioners since the introduction of the draft process analytical technology (PAT) guidance in early 2000s and its finalization in 2004 [1]. Substantial work in this area has been completed in foods and petroleum industries (and to a limited degree, the pharmaceutical industry) prior to this time; however, only a few forward looking researchers were working in the application of NIR to pharmaceutical processing prior to the FDA initiative. While the goal of the guidance is not to encumber pharmaceutical companies with additional testing burdens, achievement of the objectives of PAT often require the use of analytical solutions for on-line monitoring of pharmaceutical processes. Hammond and Warman [46] stated that greater than 70% of the on-line analytical needs of Pfizer involve NIR spectroscopy. It cannot be over-emphasized that the implementation of a NIR method (or any other on-line technique) does not constitute a PAT application; however, rapid and accurate methods for determination of critical quality attributes of a product during processing are an essential element of a successful PAT program.

The process of combining dry pharmaceutical ingredients is critical to the ultimate success of the manufacturing process. The homogeneity (or blend uniformity) achieved during the operation is of the utmost importance. Blend uniformity is typically described in terms of the mass of the final dosage unit. As the typical consumer will not consume less than an entire dosage unit, homogeneity at a scale of scrutiny less than that is frequently not considered. Recent work in NIR imaging has demonstrated some instances of this practice to be an inappropriate perspective for the overall quality of performance of a solid dosage form, and such work is covered in the analysis by NIR Imaging section of this chapter. Considerations for the mass of material sampled during blending and thus the scale of scrutiny is an important consideration when implementing NIR as a means of monitoring blending. The criteria for considering materials "homogenous" is of prime importance and must be carefully defined.

Early reports of NIR for blending was by Cuesta et al. and Wargo and Drennen [47,48] wherein, the use of NIR as a means of establishing homogeneity between samples from different locations

within a blender at a given time was demonstrated. Direct, on-line monitoring of pharmaceutical blending was demonstrated by Hailey et al. [49] and published in 1996. In this publication, the blender utilized was a V-blender with a NIR probe interfaced through windows in blender. This was the interface used again in 1998 by De Maesschalck [50]. Two differences existed between these early reports of the use of NIR for the determination of blend uniformity. The first is the basis for determining blend uniformity; the former works utilized multiple samples at a single time point; while the latter compared the same point in a blend across time. The second difference between these groups is the means by which the blend end point is calculated, although each utilized a version of either a conformity index approach (spectral matching) or a SD of some characteristic of the blend over the data collected from the blender. Numerous other reports of blend monitoring are available and employ a variety of sampling, and end-point calculation methods [51–56].

The work by Duong et al. [56] is notable as it examined the ability of NIR to monitor the uniformity of magnesium stearate in a blend. A landmark series of papers by El-Hagrasy and Drennen [57–59] demonstrated the critical nature of the location of the NIR probe in a blender. In these reports, the blending of materials in a V-blender is monitored at a number of locations around the blender. Differences in the results from the various locations are noted and an overall pattern of good inter-shell mixing and poor intra-shell mixing is observed. Lowery et al. and later Popo et al. [60,61] suggest characterization of the blend by monitoring the powder as it flows from the blender into the feed-frame of a tablet press.

An important early contribution is an estimation of the quantity of material sampled by a NIR probe during blend monitoring. This work was undertaken by Cho et al. [62] and demonstrated that the quantity of material sampled was approximately 0.15 to 0.86 g. With a typical tablet having a mass of 0.2 to 1.0 g, this quantity of material is on the correct order of magnitude to meet FDA guidelines on blend monitoring.

Wet massing of pharmaceutical ingredients, or granulation, is a commonly employed technique in pharmaceutical manufacturing. The first reports of NIR spectroscopy monitoring of this process are by Rantanen et al. [63–65] in 2000. In this work, the degree of hydration of the granules is monitored by NIR using a fiber optic probe interfaced through the side of a high-shear granulator. The authors report that a successful calibration based on four wavelengths was employed. Such a method is predicated upon the strong absorption of water in the NIR region that was used to demonstrate the ability to monitor granulation during the process.

A subsequent report by Jorgensen et al. [66] followed granulation in a similar manner, excepting that the full NIR spectrum was utilized to detect the change in moisture content of the granules and the formation of a crystalline hydrate. A fiber optic probe interfaced through the side of the granulator was used to obtain the data. This interface is common for all of the granulation monitoring applications.

In 2003, Otsuka et al. [67] reported direct determination of granule size during the granulation process by NIR spectroscopy. This was an example of directly determining the quality attribute of the product that is a function of the specific unit operation being measured. Alteration of crystalline habit of a proprietary molecule during wet granulation was reported by Li et al. [35] in 2005. Blanco et al. [36] later reported the modification of the crystalline habit of dexketoprofen trometamol during wet granulation. The work by Li and Blanco illustrates the versatility of NIR as these methods could be almost as valuable in an off-line application as they are when applied on-line. However, the ability to determine the crystalline state of the API on-line has good potential to assist formulation development by providing information on the state of the molecule as it progresses through the unit operation.

Roller compaction is another frequently employed means of mixing and consolidating pharmaceutical ingredients prior to tableting. The ability to follow this process by NIR spectroscopy was demonstrated in an off-line application by Miller [68] in 2000. Gupta et al. [69] demonstrated the ability of NIR to monitor roller compaction on-line in 2004. In this study, a fiber optic probe from a diode-array NIR was suspended above a ribbon as it exited a roller compactor. Spectra were processed to predict ribbon strength and particle size of the granules following milling. In 2005, the

ability to predict moisture content, compact density, tensile strength, and Young's modulus of roller compacted blends was demonstrated by Gupta et al. [70]. In a subsequent study, NIR was used to establish the effect of ambient moisture on the compaction behavior (during roller compaction) of microcrystalline cellulose [71]. In this application, the properties of the roller compaction ribbon were monitored in real time. It is the type of data that can be critical to control the process and consequently the success of a PAT application. Recently, Miller et al. [72] reported the application of these techniques to monitor the scale-up of a roller compaction process. This work demonstrates another important potential function of NIR in the pharmaceutical manufacturing arena: the ability to monitor a process for a given parameter at multiple scales. This capacity is not limited to roller compaction and has been demonstrated in other unit operations.

Determination of properties of film coated tablets was demonstrated by Kirsch and Drennen [73] in 1995. Here, the thickness of an ethylcellulose coating, T_{50} (50% dissolution time), and the hardness of the tablet were determined from NIR spectra measured on grating based and acousto-optic tunable filter (AOTF) spectrometers. While the performance of the grating based spectrometer was superior, the AOTF was demonstrated to perform adequately for most pharmaceutical applications and had the advantage of substantially shorter measurement times. This demonstration of the capacity for rapid tablet analysis was a critical step towards the implementation of such systems in an in-line tablet analysis system.

Further demonstrations of this potential for NIR were presented by several other researchers [52,74–78]. In 2000, Andersson et al. [79] demonstrated an in-line system for following the process of coating of tablet cores by placing a NIR-fiber optic probe inside a pan coater. It was noted that the calibration for determination of coating thickness was not only a function of the increasing signal due to the coating, but also the disappearance of signal from the core. Perez-Ramose et al. [80] utilized a similar measurement configuration and NIR monitoring of the process by a single wavelength to develop a model for the film growth on tablets.

The current literature has shown the capability NIR to monitor blending, granulation, and coating during the manufacturing process. Advances in the on-/in-line use of NIR frequently rely on characterization of in-process materials and finished dosage forms through other analytical techniques. On-line applications demonstrate the analytical potential of NIR as a sensor in pharmaceutical manufacturing; they allow the rapid chemical and physical characterization of materials during unit operations. Appropriate understanding of the product during manufacturing allows control of the production process that ultimately improves the quality of the pharmaceutical product. Briefly, NIR is a critical enabling technology for PAT.

30.6 ANALYSIS OF INTACT DOSAGE FORMS

In most cases, an entire batch of powder blend or granulation is encapsulated or tableted, then a random sample of capsules or tablets drawn and analyzed prior to final release. Failure of a sample often means rework or disposal of the entire batch. Characterization of intact dosage forms by NIR is a significant advantage because it offers the potential for on-line or at-line qualification of dosage forms. Loss of batches could be avoided, because problems could be detected and addressed immediately.

Consider the potential advantage of real-time monitoring of the tableting process. A production of a batch of 2.5 million tablets may be sold at retail for several dollars per tablet, not an unusual price today. Assume the granulation segregates in the hopper feeding a high-speed tablet press. Without NIR, it could be days before the LC data is returned, long after the production run is complete, and the lot of tablets would have to be discarded or reworked (if it is even detected). If the process had been monitored by NIR, the first unusual tablets produced as a result of segregation of the granulation would have been recognized and the process shut down, allowing the granulation to be remixed and the lot (\$7.5 million in retail value) saved.

Furthermore, a single NIR scan of one tablet, collected in a matter of seconds or even fractions of a second, may be used to qualify or quantify numerous tablet properties. NIR has shown value for monitoring drug and excipient concentrations, hardness, moisture content, dissolution rate, degradants, and so forth. Pattern recognition algorithms or other methods for qualitative analysis allow one yes/no decision to determine if a tablet meets all specifications of interest: the tablet is acceptable or the tablet is unacceptable. Given the increasing demand for correlation to a rapidly expanding list of tablet parameters, NIR is frequently a method of choice for its speed of analysis and richness of the data [1]. Because of the rapid and nondestructive nature of the method, a larger sample of a production lot can be analyzed, giving more statistical confidence to our acceptance or rejection of a lot.

The first report of NIR applied to the analysis of intact dosage forms came as a direct result of the deaths caused by cyanide-laced capsules in the early and middle 1980s. Following these poisonings, the Food and Drug Administration analyzed two million capsules by a variety of methods. In 1987, Lodder et al. [81] published a landmark paper in which intact capsules were analyzed by NIR. In this study, the quantile-BEAST cluster-analysis algorithm was used in the analysis of adulterated and unadulterated capsules.

A reproducible positioning system is critical to sampling intact capsules. To achieve this, investigators used a plastic blister glued to the center of an elliptical polished aluminum reflector. Much of the signal returned to the detector by this elliptical reflector was due to specular reflectance, revealing little about the sample. A new reflector was developed using a 90° conical reflector machined from a block of aluminum. This conical reflector minimized specular reflectance and maximized diffuse reflectance from the sample. A comparative study of the two sample configurations revealed that when the ratio of distance between spectra of dissimilar capsules to distance between spectra of similar capsules was calculated, it was nearly three times greater for the conical reflector than the elliptical reflector.

Spectra of 10 to 13 unadulterated capsules were collected at 18 wavelengths with an InfraAlyzer 400 and used as a training set. The model was tested with an equal number of unadulterated capsules. Potassium and sodium cyanide, aluminum shavings, arsenic trioxide, and other contaminants were incorporated into acetaminophen capsules and used to test the model. All the adulterated capsules were easily differentiated from unadulterated capsules, even with as few as two wavelengths.

Both capsule color and positioning of the adulterant affected the NIR analysis. The relative position of the adulterant in the capsule was predictable by the NIR. Many of the capsules studied had a white end and a colored end; the white ends of the capsules caused more light-scattering and a lower signal when oriented toward the light source, giving a first indication of the significance of sample positioning in NIR analysis of intact dosage forms.

A calibration was established for a quantile-BEAST determination of capsule KCN content. A detection limit of 2.6 mg of KCN was established in acetaminophen capsules whose average mass was 670 mg (<0.4% of the capsule weight). Nine milligrams of KCN added to an intact capsule caused the capsule to be classified as an outlier, nearly six SDs from the center of the training group.

The next paper, by Lodder and Hieftje [82], discussed NIR analysis of intact tablets. The sample cell described earlier was modified, with a smaller right circular cone at the vertex of the main cone and oriented in the opposite direction. This insert illuminated the side of the tablet away from the light source.

Commercially available aspirin tablets from two manufacturers were analyzed at 18 wavelengths using the modified sample apparatus and two other configurations. Data treatment involved principal component analysis followed by discriminant analysis with the quantile-BEAST algorithm. Cluster separation was greatest for the spectra collected with the modified sample cell and least for tablets which had been powdered and placed in a traditional sample cup. The single-reflecting cone results were slightly better than those of the powdered samples.

To further examine the utility of the double-reflecting cone, a hole was drilled in the side of a tablet and packed with 10 mg KCN. When scanned using a single reflecting cone, the sample fell into

the cluster with the training set (unadulterated) tablets. When the double-reflecting cone was used, the tablet fell well outside the unadulterated tablet cluster, revealing the value of the double-reflecting cone over the single-cone configuration.

Continuing the analysis of intact dosage forms with NIR and the quantile-BEAST algorithm, Lodder and Hieftje published an article using the technique for the quantitative and qualitative characterization of capsules with low concentrations of contaminants. Using quantile–quantile (QQ) plots, detection of subpopulations in NIR spectral clusters was possible. These subpopulations were defined as samples whose distance from the center of the training group was less than three SDs.

Ten to thirteen APAP capsules, from which the contents were removed and repacked (to minimize variation between training and test sample handling), were used for the training set and an equal number used to validate the model. Contaminated samples were produced by adding aluminum shavings (average of 208 mg per capsule) or floor sweepings (average of 221 ppm per capsule) to the capsules. Incorporating these data into QQ plots, the investigators found that the correlation coefficients for the plots containing the contaminated samples fell below the confidence level established by the QQ plots of the unadulterated training and test sets. Detection of trace contaminants was possible by this method with as few as one or two wavelengths.

In a paper published the same year, Jensen et al. [83] used NIR in the analysis of amiodarone tablets. Before spectral collection, the film-coated surfaces of the tablets were scraped off and the tablets glued to an anodized aluminum plate. In this study, the authors suggested that the interference of the film coating necessitated its removal prior to analysis. The spectrum of the pure drug was obtained to determine wavelengths of drug absorbance and was compared to the tablet spectra. Six wavelengths were chosen for the calibration. To develop the calibration, tablets ranging from 47 to 63% active ingredient were prepared in increments of 2%. The calibration provided an $r^2 = .996$ and $SEE = 0.45$.

Reproducibility of the method was determined by analyzing a group of tablets with the same concentration of active ingredient. Samples were analyzed on a variety of sample backings and at three storage conditions (room temperature, 40°C, and room temperature with storage in a desiccator). Although the results were slightly less variable for samples stored at 40°C, NIR prediction for samples stored under all three conditions was virtually identical. Subsequent reports demonstrated the lack of necessity of removal of a tablet film coat prior to NIR analysis. An early demonstration of analysis through a tablet coating was reported by Kirsch and Drennen [73].

An investigation into the determination of degradation products by NIR was published in 1990 by Drennen and Lodder [84]. The major degradation process in aspirin tablets is the hydrolysis of aspirin to salicylic acid. One of two USP methods must be performed to verify tablet aspirin content and both are time-consuming. A second analysis by HPLC must be performed to verify that salicylic acid levels do not exceed 0.3% of the tablet mass.

In this study, tablets were stored in a hydrator for up to 168 h with tablets withdrawn at regular intervals. After removal from the hydrator, the tablets were weighed and NIR spectra collected prior to the HPLC analysis. Spectra of the intact tablets were collected on an InfraAlyzer 500 in the 1100 to 2500 nm region, using the double-reflecting sample apparatus described by Lodder and Hieftje [82]. The spectra were processed by principal component analysis, and the scores analyzed by the quantile-BEAST algorithm.

The study had three objectives: first, changes in NIR spectra were correlated to the time aspirin tablets spent in the hydrator (the calibration had a correlation coefficient of 0.95 and SEE of 18.8 h); second, a calibration was developed for the prediction of tablet salicylic acid content (the researchers ensured that prediction of salicylic acid was based on changes in salicylic acid concentration, and not some related process, such as absorption of moisture, by evaluation of loadings spectra from principal component analysis of the data. The HPLC-determined salicylic acid levels ranged from 0.36 to 1.66 mg, and the NIR method allowed prediction of the degradant with a standard error of 144 mg); and third, the mass of water absorbed by the tablets was determined by NIR spectroscopy.

Given the strong absorbance of water in the NIR region, a correlation between NIR spectra and water absorbed is not surprising. However, the amount of water absorbed by the tablets, determined by weighing the tablets on an analytical balance, was <2.5 mg for all samples. In this study, the amount of water absorbed was predicted with a standard error of $163\ \mu\text{g}$. Thus, even very small changes in dosage form moisture content can be detected by this method.

Two book chapters discussing NIR analysis of tablets were published in 1991. The first was from work presented at the 4th International Conference on Near Infrared Spectroscopy [85]. In this paper, Stark used a newly developed diode array spectrometer which scanned the 520 to 1800 nm region, useful in the analysis of macro- and micro-specimens. Samples were placed on a glass slide prior to analysis, maintained at an angle to the probe. With this configuration, light reflected by the sample was detected, but light reflected by the glass was not. Spectra were collected from a 1 mg sample of acetaminophen powder. The sample was subsequently divided into 500, 250, and 125 mg; reasonable spectra were collected from all these samples. Diffuse reflectance spectra of intact acetaminophen, ibuprofen, and antacid tablets were successfully collected. No comparative studies or quantitative or qualitative analyses of the spectra were performed.

Monfre and DeThomas [86] published a chapter describing a NIR calibration for QC monitoring of a marketed vasodilator. The NIR analysis was performed on a NIRSystems Model 6500, using an aperture plate to assist in tablet positioning. For the calibration, individual tablets were crushed to a fine powder and scanned. Second derivative spectra were used in the analysis to minimize the light scattering differences between the samples. A bias correction was introduced to factor in the scattering efficiency of the tablets vs. the powders. Because the excipient concentrations were not constant, a drug absorbance wavelength was ratioed by an absorbance wavelength primarily due to the formulation matrix, and these normalized values were then used for the calibration.

HPLC was used as the reference method and the tablets were found to vary between 96 and 102% of labeled strength. To determine the accuracy and precision of the method, one tablet was analyzed ten times with sample removal and replacement between scans. The NIR-determined value was within 0.5 mg of the HPLC-determined value, indicating the accuracy of the method. Tablet placement on the spectrometer played an important role in the precision of the method, even with the use of the aperture plate.

An increasing rate of publications on the determination of properties of finished dosage forms followed this initial work; between 1995 and 2005 more than 100 papers were published on this subject. Early work was concerned with the comparison of transmission and reflectance. An example of this is found in Gottfries et al. [87] from 1996, where transmission measurements were found to have a lower RMSEP (1.06 vs. 2.83) than reflection measurements for the determination of metoprolol succinate. Merckle and Kovar [88] compared transmission and reflection measurements of acetylsalicylic acid formulations and found both performed adequately without a significant difference between them. A comparison by Thosar et al. [89] indicated an advantage in using transmission, but noted that both methods demonstrated performance that was suitable to task. Cogdill et al. [90] indicated that while both reflection and transmission were suitable for determination of API in tablets in a system designed for in-line application, reflection was demonstrated to be much less sensitive to sample position.

The late 1990s represented a significant maturation of the use of NIR for tablet analysis, a summary of the analytes, ranges and NIR mode used is found in Table 30.1. Many researchers published accounts of successfully calibrating for tablet properties by NIR analysis [89,91–108]. Ebube et al. [109] reported an analysis of magnesium stearate at concentrations down to 0.25% in compacts of microcrystalline cellulose. Gustafsson et al. [110] reported the use of IR and NIR for the determination of compact density, particle shape, tablet axial tensile strength, and drug release characteristics. This work is an example of combining multiple spectral ranges (IR and NIR) and process data (compaction behavior) to create comprehensive models for tablet performance (ideas well aligned with PAT).

TABLE 30.1**Reported NIR Calibrations of Pharmaceutical Analytes and the Associated Sampling Method for Each Method**

Analyte	Range (or lowest level studied, % w/w)	Sampling	Reference
Potassium cyanide (in acetaminophen capsules)	0.4	Reflection with a conical reflector accessory	[81]
Potassium cyanide (in aspirin tablets)	2	Reflection with a double reflecting conical accessory	[82]
Aluminum shavings	31	Reflection with a double reflecting conical accessory	[114]
Floor sweepings	33	Reflection with a double reflecting conical accessory	[114]
Amiodarone	47–63	Reflection (tablet coating removed)	[83]
Salicylic acid in Aspirin tablets	0.07–0.3	Reflection with a double reflecting conical accessory	[84]
Metoprolol	20–25	Transmission and reflection	[87]
Aspirin	7.0–21	Transmission and reflection	[88]
Paracetamol	76–93	Reflection	[91]
Magnesium stearate	0.25–2.0	Reflection	[109]
Paracetamol	76–93	Reflection	[111]
Gemfibrozil	67–89	Reflection	[93]
Steroid (Proprietary)	2.9–18	Transmission	[94]
Theophylline	1–40	Transmission and reflection	[89]
Ibuprofen	49–90	Transmission	[95]
Caffeine	13.7	Reflection	[98]
Water	1.74–5.32	Transmission	[99]
Sulfamethazine	60	Reflection	[100]
Bromazepam	0.60–3.9	Reflection	[101]
Clonazepam	1.4–2.6	Reflection	[101]
Paracetamol	45	Reflection	[102]
Amantadine hydrochloride	17	Reflection	[102]
Cimetidine	66–86	Reflection	[103]
Aminopyrine	28	Reflection	[104]
Phenacetin	28	Reflection	[104]
Roxithromycin	19.5	Reflection	[105]
Erythromycin	28.1	Reflection	[105]
Riboflavin	0.41–2.3	Transmission	[106]
Ibuprofen	0.7	Transmission	[107]
Mirtazapine	5.5–14.5	Reflection	[108]

Research through the 1990s dealt primarily with demonstration of the ability of NIR to determine different tablet properties in a regulatory environment (validation). An account of a NIR analytical methods validation by Moffat et al. [111] was published in 2000. The target of this investigation was to demonstrate that a NIR analytical method met the criteria established by ICH Q2 [112,113]. The accuracy, precision, specificity, detection limit, quantification limit, linearity, range, robustness, and system suitability testing were demonstrated to be appropriate for use in routine testing. While this type of validation is based upon the needs of other analytical techniques, it served as an example of the validation of a NIR-based method for tablet analysis. Other researchers followed in validating NIR analyses of tablets for use in release testing [95–98].

Prediction of the dissolution rate of tablets is another area of application of NIR spectroscopy. Carbamazepine is used for the treatment of epilepsy and consistent dissolution of the dosage form is critical in the maintenance of therapeutic blood levels of drug. In a 1991 study by Zannikos et al. [115], dissolution profiles of brand-name and generic carbamazepine tablets were compared after storage in conditions of high humidity. The calibration was based upon the percentage of drug in solution after 1 h in a USP Dissolution Apparatus II.

Spectra of intact tablets were collected on an InfraAlyzer 500, after which the dissolution rates of the tablets were determined. The NIR spectra and dissolution profiles of the generic tablets changed little during the 5 days of high-humidity storage, but the spectra and dissolution profiles of the brand-name tablets changed significantly. A calibration based upon principal component analysis followed by the bootstrap algorithm was developed for the brand-name tablets. The R was .99, and the SEP for extent of dissolution after 1 h was 6.8%.

Further storage of the tablets did not affect the dissolution profiles or the NIR spectra of the tablets appreciably, but after 5 days other absorbance peaks began to appear in the NIR spectra, which were attributed to degradation products in the tablet. The authors surmised that the peaks could have resulted from a chemical or physical change that altered the dissolution rate.

Drennen and Lodder [116] published a paper in 1991 comparing the performance of the improved quantile-BEAST algorithm with that of the Mahalanobis distance in the qualitative analysis of carbamazepine tablets. While the Mahalanobis distance calculation assumes that spectral variations associated with both the calibration and test set are random, in complex pharmaceutical mixtures this may not be the case. The bootstrap algorithm, on the other hand, is a nonparametric test which can be used with nearly any spectral data distribution.

In this study, the dissolution profiles of carbamazepine tablets exposed to conditions of high humidity were classified according to the Mahalanobis distance calculation and the bootstrap method using both full (701 wavelength) and principal component spectra collected on an InfraAlyzer 500. This was the first report of the use of full spectra and required substantial computing power relative to readily accessible systems of the time. In multiple tests, the bootstrap calculation was shown to provide more accurate qualitative results than the Mahalanobis calculation. In one experiment, nine tablets with a slow dissolution rate were used as a training set. Twenty-one tablets with a variety of dissolution rates were used to test the model. The modified bootstrap calculation correctly identified all tablets with a faster dissolution rate than the training set, while the Mahalanobis calculation incorrectly identified 58% of the tablets with a higher dissolution rate. The quantile-BEAST algorithm gave better precision, accuracy, and speed than the Mahalanobis calculation in nearly all cases.

Recent reports of prediction of dissolution behavior are focused on prediction of specific, release testing criteria. Tatavarti Aditya et al. [100] have reported prediction of the quantity of drug released at 120 min (Q120). Donoso and Ghaly [117] and Freitas et al. [118] developed models to predict the behavior of the tablet in specific pH buffer at a specific time. Each of these authors developed individual calibration for the quantity of drug released at several times for each of 3 pH conditions for dissolution testing. Donoso [119] also reported using NIR as a means of predicting the disintegration time for a tablet.

30.7 HARDNESS

In a 1993 review of pharmaceutical applications of NIR spectroscopy, Drennen and Lodder [120] presented new research on the prediction of tablet hardness based upon NIR spectral changes. Tablets ranging in hardness from 0.3 to 10.75 kilopons (kp) were analyzed nondestructively by NIR, then subjected to the destructive reference test. Increasing tablet hardness was found to cause an upward shift in the spectral baselines, probably due to a reduction of light scattering from the tablet. It was surmised that a harder and smoother tablet surface reduced the light scattering from the surface, allowing more light to penetrate the sample and causing increased absorbance. Tablets were easily classified according to hardness using the quantile-BEAST algorithm. Spectral changes were found

to correlate well with variations in hardness. Prediction of hardness provided SEE and SEP values of approximately 0.6 kp.

Following the initial example of the use of NIR to predict hardness, researchers have used NIR to predict hardness [121] and related properties such as tensile strength of tablets [100,110], or have demonstrated the effect of compression force on tablets [108,109].

The use of NIR in the qualification of clinical batches of tablets is an application for which this technique is well suited. When medications are dispensed for use in clinical trials, the blister packs in which they are distributed often contain tablets with a range of doses. The tablets are usually identical in appearance, making visual classification nearly impossible. Verification of the correct tablet configuration in the packaging by a noninvasive and nondestructive method would be a significant advantage in the quality control inspection of such packages.

In the first study published on this tablet analysis application, Dempster et al. [122] used three sampling configurations to investigate the classification of an experimental drug present in tablets in 2, 5, 10, and 20% concentrations, a matching placebo, and a marketed drug used as a clinical comparator. The first method of tablet analysis required the removal of the tablets from the blister packs and scanning them directly through the spectrometer window. In the second approach, the tablets were scanned through the plastic packaging using the spectrometer window. With the third arrangement, the tablets were analyzed through the plastic blister packaging with a fiber-optic probe.

A NIRSystems Model 6500 was employed in the analysis with a ceramic disk used as the reference. Second-derivative spectra were used in the data analysis. The identification and classification algorithms used were supplied by the instrument vendor. In the first configuration, all but the 2% tablets were easily classified. The 2% tablets were not be differentiated from the placebo. Using the second and third configurations, only the 10 and 20%, placebo and clinical comparator tablets could be properly classified.

This lack of ability to identify and classify the range of dosages in the clinical batch indicates the need for further research in this area. Substituting another plastic for the white opaque blister packaging used may have improved the results of the analysis, since this packaging would be expected to be an excellent light scatterer, decreasing the signal of the tablet within the blister packaging.

The second application of NIR spectroscopy in the analysis of intact tablets from clinical batches was published in a 1994 paper by Aldridge et al. [123]. A NIRSystems Model 6500 with a custom sampling configuration was used for spectral collection of the blister-packed samples, and the second derivative spectra were used in the analysis. SpectralonTM was used for reference. Although certain peaks in the NIR spectra were attributed to the hydrocarbon functionality of the packaging material, the spectral features of the tablets within were clearly visible.

The Mahalanobis distance calculation was used for discriminant analysis. These distances were plotted in a control chart, revealing that several samples were in danger of being misclassified based upon the Mahalanobis distance calculation alone. When the residual ratios of the spectra were calculated and plotted in a control chart, however, the probability of misclassifying a sample was greatly diminished.

In a paper by Kirsch and Drennen [73], intact theophylline tablets coated with an ethylcellulose polymer were analyzed by grating-based and AOTF spectrometers. The purpose of the work was threefold. First, tablets were coated with increasing levels of ethylcellulose to vary the dissolution profiles of the dosage forms. After NIR spectra were collected, the tablets were subjected to dissolution in a USPDA II. The time required for 50% of the drug to enter solution was used as the measure of dissolution rate. Principal component regression was used to develop the calibration. The calibration provided a SEE of 2.8 min, a coefficient of determination of .977, and an SEP of 6.6 min, with time to 50% dissolution values ranging from 48 to 93 min.

Second, the potential of this method in the monitoring of the film-coating process was investigated. Tablets coated with 2 to 7% ethylcellulose were used to determine the utility of this method in the prediction of film coat thickness. NIR spectral changes were found to correlate well with film thickness. Using the first principal component, a SEE of 0.0002 in. was obtained, for coating

thicknesses ranging from 0.001 to 0.003 in. The use of NIR spectroscopy as a means of on-line evaluation of film-coating is currently under investigation in this laboratory.

The third experiment was a NIR-based determination of the hardness of coated tablets. Thirty-eight tablets whose hardness ranged from 6 to 12 kp were first analyzed spectroscopically, and then by the destructive reference method. Prediction of hardness with a standard error of 0.6 kp was possible even after removal of the baseline shifting with MSC. This research confirmed the results of hardness studies reported earlier by Drennen and Lodder [120]. Further research is being conducted to characterize this phenomenon.

In another paper by Lodder et al. [124], the qualification of a number of tablet characteristics was performed in a comparative study of two classification algorithms: soft independent modeling of class analogies (SIMCA) and the quantile-BEAST. The study involved qualitative classification of tablet hardness, moisture content, dissolution rate, and degradant concentration.

An evaluation of the performance of these algorithms in predictions using inside model space and outside model space was conducted. In principal component regression, principal axes highly correlated with sample constituents of interest are considered to be inside model space, while axes typically attributed to spectral noise are termed outside model space.

SIMCA provided highly variable results, occasionally offering optimum performance with outside model space, while the quantile-BEAST gave better results overall and more consistent prediction. The best results were obtained when the quantile-BEAST algorithm was used with full spectra, with no principal axis transformation prior to analysis.

In a later work, Cogdill, Anderson, and Drennen [125] demonstrated the use of Hotelling's T^2 and spectral residuals (Q_{res}), and proposed wavelength uncertainty and spectral noise tests to demonstrate the suitability of the measurement for prediction of a concentration. Hotelling's T^2 and Q_{res} were used to verify the overall suitability of the prediction of the API concentration. Factors influencing these two metrics include the interaction between the model and sample changes, sample interface changes, and spectrometer performance. The remaining two metrics were utilized as indicators of spectrometer performance and were not related to the model used to predict API concentration or hardness.

30.8 CONSIDERATIONS FOR INTACT DOSAGE FORM ANALYSIS

Numerous styles and brands of instruments and sample cells have been used for the analysis of tablets. The authors currently use several brands of instrumentation for tablet analysis, including filter based, diffraction-grating based, and AOTF-based instrumentation. Detector configurations for tablet analysis are evolving slowly towards an optimum design; however, the standard designs offered by most instrument manufacturers are suitable. Tablets have been successfully analyzed with integrating spheres and with a standard dual-angled detector configuration. Intact tablets are analyzed in both diffuse reflectance and transmission modes.

The first analyses of individual intact tablets and capsules involved the use of reflective aluminum sample cells, designed specifically for tablets [82] or capsules [81], that allowed illumination of all sample surfaces. Illumination of all sample surfaces has proven to offer enhanced sensitivity. The authors now prefer the latest configuration of the original sample cell, the CAPCELL™ (Optical Prototypes, Inc., Natrona Heights, PA 15065), for analysis of individual intact tablets and capsules.

Sample positioning variability is the single largest source of error in NIR analysis of tablets, regardless of whether diffuse reflectance or transmission measurements are used. Hardware, methodology, and mathematics may be used to reduce this error. Tablet-specific sample cells that allow consistent positioning of tablets are valuable in reducing this error. A method involving the collection of three spectra per tablet, with 120° rotation of the sample or the sample cell/sample combination between spectra, is used by the authors. Using the mean (or median) spectrum for each tablet

significantly reduces the spectral variability by averaging out the positioning effects. The median calculation results in less weighting by particularly odd spectra than does the mean.

The individual tablet spectra must then always be corrected for baseline shifting prior to analysis. Many techniques have been attempted, but the second derivative and MSC calculations are most common. This baseline correction is critical even if an average tablet spectrum is used.

Curved surfaces, debossing, and scoring are factors that affect the spectrum of a tablet as positioning is varied, but the spectral effects of such factors can be reduced by the methods just discussed. Natural variations in tablet mass and hardness will affect a tablet's spectrum, primarily through spectral baseline shifting. Some work by Baxter [126] involved a unique method of normalizing tablet weight variations.

Baxter concluded that because NIR spectra are "in essence a picture of active per unit area" and do not allow detection of differences in tablet weight, reference assay values should be normalized for tablet weight, multiplying the HPLC reference value by the theoretical tablet weight and dividing by the actual tablet weight. Values predicted from this calibration must then be denormalized by multiplying the NIR predicted value by the actual tablet weight, divided by the theoretical weight. Baxter observed a reduction in residual values from 2.17 to 1.57% for 228 tablets for which active ingredient concentrations were predicted.

In 1997, Candolfi et al. [41] studied the sources of variance in NIR measurements of tablets and capsules. The sources of variation studied were measurement repeatability, sample positioning, day-to-day variability, object-to-object variability within a batch, and batch-to-batch variability. For tablets, positioning and time between measurements were the least significant factors; for capsules, sample positioning and time contributed the most to the variability of the spectra. Further remarks indicate that the researchers doubt that the contents of the gelatin capsules were probed. This work gave one of the early indications of the importance of sample positioning that remains a central issue to the present.

Researchers validating or performing tests claiming to validate analytical methods based upon the chemometric analysis of NIR data must take care to establish the complete independence of the validation samples. Take, for example, a data set of eight spectra collected from each of 20 tablets sampled from 10 batches. Setting aside one of the eight spectra collected from each tablet does not constitute an independent validation of the method; similarly, setting aside half of the tablets from each batch does not constitute an independent validation set. Even though the number of batches is limited, the only independent validation set available is the removal of a number of batches from the training and testing set to the validation set. From a compliance perspective, once a batch of samples have been used to validate the method, changes to the method (particularly those based upon the results of testing the validation set) render those samples not longer suitable for validation. In this instance, testing to demonstrate validation will require a new set of samples from which the data must be collected.

30.9 NEAR-INFRARED IMAGING

Hyperspectral imaging (or NIR imaging) increases the information density of NIR spectroscopy by combining spectra with spatial information of the locations from which the spectra originate. NIR as a single-point measurement has been demonstrated to be a robust technique for analysis of pharmaceutical ingredients; however, it lacks any spatial information beyond the volume of the sample represented by the spectrum. In hyperspectral imaging, the spatial domain is much like the field of view for a photograph; the difference is that instead of a color or intensity at each point on the image an entire NIR spectrum is captured for that point (or pixel). NIR imaging spectroscopy is not a new technique application and examples have been reported in the mid-1990s [127,128], however, it has found new interest in the wake of the FDA's PAT initiative [1].

The role of hyperspectral imaging in pharmaceutical manufacturing was discussed by Clarke et al. [129–131]. In these reports, it is suggested that the combination of chemical and spatial information provided by NIR imaging can greatly enhance the knowledge of the product and the process that has produced it. A subsequent paper by Clarke et al. [132] described the depth of penetration in terms of the thickness of penetration at which 50% of the substrate signal has been attenuated by the layers of cellulose placed on top of it. The values for this condition varied with wavelength and per sample. At 2380 nm, the 50% attenuation was approximately 27 μm ; at 1675 nm, 39 to 61 μm ; and at 1100 nm, ca. 180 μm .

Lyon et al. [133] reported on the value of NIR imaging for trouble-shooting formulations and pharmaceutical processes by careful analysis of the images. In this work, the importance of image analysis techniques is emphasized as a means of utilizing the data for purposes beyond intuitive visual inspection. The use of histograms to describe populations within an image is a key tool to gain maximum knowledge from images. Further discussions of the application of imaging to solid-dosage processes [8,134]; wherein LaPlant and Lodder focused on the data processing requirements of imaging spectroscopy.

Hyperspectral imaging of freeze-dried solid-dosage formulations have been recently published [135]. The ability to characterize both the morphological and chemical properties of a freeze-dried formulation is a critical element to the complete description the dosage forms studied.

30.10 CONCLUSIONS

Tremendous advances have been made recently in the use of NIR spectroscopy for the analysis of pharmaceutical dosage forms. Just 25 years ago, NIR spectroscopy was used in a way that offered relatively few advantages over other analytical methods for the analysis of dosage form drug content, requiring extractions with organic solvents prior to sample analysis. With advances in instrumentation, software, and sample handling, rapid characterization of intact dosage forms has become a reality. The pharmaceutical industry is beginning to implement NIR methods to monitor many phases of the manufacturing process, from the arrival of bulk raw material at the loading dock, to the inspection of tablets for final release.

Myths about the “black box” nature of this method have been debunked, and as those involved in analytical methods development, process control, and quality assurance acquire a more thorough understanding of NIR spectroscopy and its capabilities, pharmaceutical applications will become even more widespread. NIR instruments are becoming faster, smaller, and less expensive, increasing their potential for application as process monitors in many phases of the manufacturing process. Similarly, increased computer storage and power are becoming rapidly cheaper. Pharmaceutical manufacturers are under increasing pressure to validate their processes and to provide extensive documentation of ongoing validation activities. NIR has proven itself to be a rapid and rugged analytical method capable of continuous on-line process monitoring, making it a valuable method to couple with ongoing validation activities.

The FDA has recognized the value of NIR spectroscopy, and has already approved the method for the analysis of lincomycin in a veterinary product [136]. More recently, NIR spectroscopy has been employed in the qualification of incoming raw materials, and the FDA is working with pharmaceutical manufacturers to develop and implement other NIR-based methods, with several having been approved.

In many ways, NIR spectroscopy is an ideal method for pharmaceutical process control, particularly for the analysis of intact dosage forms. As production costs, including analytical expenses, continue to increase, the advantages of NIR spectroscopy will become more attractive. With NIR spectroscopy, the pharmaceutical industry will move one step closer to “zero-defect” quality control, making the costs associated with the method’s development well spent.

REFERENCES

1. FDA, PAT-A Framework for Innovative Manufacturing and Quality Assurance, <http://fda.gov/cder/OPS/PAT.htm>
2. Ciurczak, E.W., Uses of near-infrared in pharmaceutical analysis. *Appl. Spectrosc. Rev.*, **23**: 147–163 (1987).
3. Martin, K.A., Recent advances in near-infrared reflectance spectroscopy. *Appl. Spectrosc. Rev.*, **27**: 325–383 (1992).
4. Drennen, J.K., E.G. Kraemer, and R.A. Lodder, Advances and perspectives in near-infrared spectrophotometry. *Crit. Rev. Anal. Chem.*, **22**: 443–475 (1991).
5. Kirsch, J.D. and J.K. Drennen, Near-infrared spectroscopy: Applications in the analysis of tablets and solid pharmaceutical dosage forms. *Appl. Spectrosc. Rev.*, **30**: 139–174 (1995).
6. Ciurczak, E.W. and J.K. Drennen, Pharmaceutical applications of near-infrared spectroscopy, in *Near-infrared applications in biotechnology*, edited by R. Raghavachari. Boca Raton, FL: CRC Press, pp. 349–366 (2001).
7. Cogdill, R.P., C.A. Anderson, and J.K. Drennen, III, Using NIR spectroscopy as an integrated PAT tool. *Spectroscopy (Duluth, MN, United States)*, **19**: 104–109 (2004).
8. LaPlant, F., Factors affecting NIR chemical images of solid dosage forms. *Am. Pharmaceutic. Rev.*, **7**: 16, 18–24 (2004).
9. Patonay, G., *Advances in near-infrared measurements*. Greenwich: JAI Press (1993).
10. Osborne, B.G. and T. Fearn, *Near-infrared spectroscopy in food analysis*. New York: Longman Scientific and Technical (1986).
11. Hrushchka, W.R., *Near-infrared technology in the agricultural and food industries*, eds. P. Williams and K. Norris. American Association of Cereal Chemists St. Paul, MN (1987).
12. Burns, D.A. and E.W. Ciurczak, *Handbook of near-infrared analysis*. New York: Marcel Dekker (1992).
13. Ozaki, Y., et al., *Near-infrared spectroscopy: Principles, instruments, applications*. Weinheim (Germany): Wiley-VCH (2002).
14. McClure, F., *Near-infrared spectrometry: Learning the fundamentals*. VCH, US (2006).
15. Ciurczak, E.W. and J.K. Drennen, III, *Pharmaceutical and medical applications of near-infrared spectroscopy. Practical spectroscopy*. New York: Marcel Dekker, p. 192 (2002).
16. Raghavachari, R., *Near-infrared applications in biotechnology (Practical spectroscopy)*. Boca Raton, FL: CRC Press (2001).
17. Corti, P., E. Dreassi, and S. Lonardi, Near infrared reflectance analysis: Features and applications in pharmaceutical and biomedical analysis. *Il Farmaco*, **48**: 3–20 (1993).
18. Josefson, M., et al., *New opportunities with NIR spectrometry in the analysis of dosage forms*. *Eur. J. Pharm. Sci.*, **2**: 82–83 (1994).
19. Plugge, W. and C.V.d. Vlies, The use of near infrared spectroscopy in the quality control laboratory of the pharmaceutical industry. *J. Pharm. Biomed. Anal.*, **10**: 797–803 (1992).
20. MacDonald, B.F. and K.A. Prebble, Some applications of near-infrared reflectance analysis in the pharmaceutical industry. *J. Pharm. Biomed. Anal.*, **11**: 1077–1085 (1993).
21. Sinsheimer, J.E. and A.M. Keuhnelian, Near-infrared spectroscopy of amine salts. *J. Pharm. Sci.*, **55**: 1240–1244 (1966).
22. Oi, N. and E. Inaba, Determination of allylisopropylacetureide and phenacetin in pharmaceutical preparations by near infrared absorption spectroscopy. *Yakugaku Zasshi*, **87**: 213–215 (1967).
23. Sinsheimer, J.E. and N.M. Poswalk, Pharmaceutical applications of the near infrared determination of water. *J. Pharm. Sci.*, **57**: 2006–2010 (1968).
24. Sherken, S., Rapid near-infrared spectrophotometric method for determination of meprobamate in meprobamate tablets. *J. Ass. Offic. Anal. Chem.*, **51**: 616–618 (1968).
25. Allen, L., Quantitative determination of carisoprodol, phenacetin, and caffeine in tablets by near-IR spectrometry and their identification by TLC. *J. Pharm. Sci.*, **63**: 912–916 (1974).
26. Zappala, A.F. and A. Post, Rapid near IR spectrophotometric determination of meprobamate in pharmaceutical preparations. *J. Pharm. Sci.*, **66**: 292–293 (1977).
27. Corti, P., et al., Application of near infrared reflectance spectroscopy to pharmaceutical control: I. Preliminary investigation of the uniformity of tablets content. *Analysis*, **18**: 112–116 (1990).

28. Becconsall, J.K., J. Percy, and R.F. Reid, Quantitative photoacoustic spectroscopy of propranolol/magnesium carbonate powder mixtures in the ultraviolet and near-infrared regions. *Anal. Chem.*, **53**: 2037–2040 (1981).
29. Ciurczak, E.W. and T.A. Maldacker, Identification of actives in multicomponent pharmaceutical dosage forms using near-infrared reflectance analysis. *Spectroscopy*, **1**: 36–39 (1986).
30. Ciurczak, E.W. and R.P. Torlini, Analysis of solid and liquid dosage forms using near-infrared reflectance spectroscopy. *Spectroscopy*, **2**: 41–43 (1987).
31. Chasseur, J.C., On-line assay of cimetidine granules by near infrared reflectance spectroscopy. *Chim. Oggi*, **6**: 21–24 (1987).
32. Osborne, B.G., Determination of nicotinamide in pre-mixes by near-infrared reflectance spectrometry. *Analyst*, **112**: 313–315 (1987).
33. Lodder, R.A. and G.M. Hieftje, Quantile beast attacks the false-sample problem in near-infrared reflectance analysis. *Appl. Spectrosc.*, **42**: 1351–1365 (1988).
34. Gimet, R. and T. Luong, Quantitative determination of polymorphic forms in a formulation matrix using the near infrared reflectance analysis technique. *J. Pharm. Biomed. Anal.*, **5**: 205–211 (1987).
35. Li, W., et al., Determination of polymorph conversion of an active pharmaceutical ingredient in wet granulation using NIR calibration models generated from the premix blends. *J. Pharm. Sci.*, **94**: 2800–2806 (2005).
36. Blanco, M., et al., Near infrared spectroscopy in the study of polymorphic transformations. *Anal. Chim. Acta*, **567**: 262–268 (2006).
37. Bauer, J.F., W. Dziki, and J.E. Quick, Role of an isomorphic desolvate in dissolution failures of an erythromycin tablet formulation. *J. Pharm. Sci.*, **88**: 1222–1227 (1999).
38. Corti, P., et al., Application of NIRS to the quality control of pharmaceuticals: Ketoprofen assay in different pharmaceutical formulae. *Pharm. Acta Helv.*, **64**: 140–145 (1989).
39. Corti, P., et al., Application of near-infrared reflectance to the analytical control of pharmaceuticals. Assay of ranitidine hydrochloride and water content in tablets. *Pharm. Acta Helv.*, **65**: 28–32 (1990).
40. Ryan, J.A., et al., Rapid verification and identity and content of drug formulations using mid-infrared spectroscopy. *J. Pharm. Biomed. Anal.*, **9**: 303–310 (1991).
41. Candolfi, A., D.L. Massart, and S. Heuerding, Investigation of sources of variance which contribute to NIR-spectroscopic measurement of pharmaceutical formulations. *Anal. Chim. Acta*, **345**: 185–196 (1997).
42. Candolfi, A., et al., Comparison of classification approaches applied to NIR-spectra of clinical study lots. *J. Pharm. Biomed. Anal.*, **16**: 1329–1347 (1998).
43. De Maesschalck, R. and T. Van den Kerkhof, Implementation of a simple semi-quantitative near-infrared method for the classification of clinical trial tablets. *J. Pharm. Biomed. Anal.*, **37**: 109–114 (2005).
44. Corti, P., et al., Application of near-infrared reflectance spectroscopy (NIRS) to several antibiotic compounds. *Process Control Qual.*, **2**: 131–142 (1992).
45. Blanco, M., et al., Determination of ascorbic acid in pharmaceutical preparations by near infrared reflectance spectroscopy. *Talanta*, **40**: 1671–1676 (1993).
46. Hammond, S. and M. Warman, *The use of process analytical technology in Pfizer manufacturing plants*. Abstracts of Papers, 225th ACS National Meeting, New Orleans, LA, United States, March 23–27, p. ANYL-276 (2003).
47. Cuesta Sanchez, F., et al., Monitoring powder blending by NIR spectroscopy. *Fresenius J. Anal. Chem.*, **352**: 771–778 (1995).
48. Wargo, D.J. and J.K. Drennen, Near-infrared spectroscopic characterization of pharmaceutical powder blends. *J. Pharm. Biomed. Anal.*, **14**: 1415–1423 (1996).
49. Hailey, P.A., et al., Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy. Part I. System development and control. *J. Pharm. Biomed. Anal.*, **14**: 551–559 (1996).
50. De Maesschalck, R., et al., Online monitoring of powder blending with near-infrared spectroscopy. *Appl. Spectrosc.*, **52**: 725–731 (1998).
51. Sekulic, S.S., et al., Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy. Part II. Qualitative approaches to blend evaluation. *J. Pharm. Biomed. Anal.*, **17**: 1285–1309 (1998).

52. Blanco, M., et al., Analytical control of pharmaceutical production steps by near infrared reflectance spectroscopy. *Anal. Chim. Acta*, **392**: 237–246 (1999).
53. Blanco, M., et al., Development and validation of a near infrared method for the analytical control of a pharmaceutical preparation in three steps of the manufacturing process. *Fresenius J. Anal. Chem.*, **368**: 534–539 (2000).
54. Ufret, C. and K. Morris, Modeling of powder blending using on-line near-infrared measurements. *Drug Dev. Ind. Pharm.*, **27**: 719–729 (2001).
55. Blanco, M., R. Gozalez Bano, and E. Bertran, Monitoring powder blending in pharmaceutical processes by use of near infrared spectroscopy. *Talanta*, **56**: 203–212 (2002).
56. Duong, N.-H., et al., A homogeneity study using NIR spectroscopy: Tracking magnesium stearate in bohle bin-blender. *Drug Dev. Ind. Pharm.*, **29**: 679–687 (2003).
57. El-Hagrasy Arwa, S., F. D’Amico, and K. Drennen James, III, A Process Analytical Technology approach to near-infrared process control of pharmaceutical powder blending. Part I: D-optimal design for characterization of powder mixing and preliminary spectral data evaluation. *J. Pharm. Sci.*, **95**: 392–406 (2006).
58. El-Hagrasy Arwa, S., M. Delgado-Lopez, and K. Drennen James, III, A Process Analytical Technology approach to near-infrared process control of pharmaceutical powder blending. Part II: Qualitative near-infrared models for prediction of blend homogeneity. *J. Pharm. Sci.*, **95**: 407–421 (2006).
59. El-Hagrasy Arwa, S. and K. Drennen James, III, A Process Analytical Technology approach to near-infrared process control of pharmaceutical powder blending. Part III: Quantitative near-infrared calibration for prediction of blend homogeneity and characterization of powder mixing kinetics. *J. Pharm. Sci.*, **95**: 422–434 (2006).
60. Lowery, M., et al., An examination of dynamic and static near infrared measurements of pharmaceutical blends. *Near Infrared Spectroscopy, Proceedings of the International Conference*, 9th, Verona, Italy, June 13–18, pp. 467–473 (1999), (2000).
61. Popo, M., et al., Blend uniformity analysis using stream sampling and near infrared spectroscopy. *AAPS PharmSciTech* [electronic resource], **3**: E24 (2002).
62. Cho, J., et al., Effective mass sampled by NIR fiber-optic reflectance probes in blending processes. *Anal. Chim. Acta*, **348**: 303–310 (1997).
63. Rantanen, J., et al., Next generation fluidized bed granulator automation. *AAPS PharmSciTech* [electronic resource], **1**: E10 (2000).
64. Rantanen, J., et al., Use of the near-infrared reflectance method for measurement of moisture content during granulation. *Pharm. Dev. Technol.*, **5**: 209–217 (2000).
65. Rantanen, J., et al., In-line moisture measurement during granulation with a four-wavelength near infrared sensor: An evaluation of particle size and binder effects. *Eur. J. Pharm. Biopharm.*, **50**: 271–276 (2000).
66. Jorgensen, A., et al., Hydrate formation during wet granulation studied by spectroscopic methods and multivariate analysis. *Pharm. Res.*, **19**: 1285–1291 (2002).
67. Otsuka, M., Y. Mouri, and Y. Matsuda, Chemometric evaluation of pharmaceutical properties of antipyrene granules by near-infrared spectroscopy. *AAPS PharmSciTech*, **4**: 375–381 (2003).
68. Miller, R.W., The use of near infrared technology to map roller compaction processing applications. *Proceedings — Institute for Briquetting and Agglomeration, Biennial Conference*, San Diego, California, **26**: pp. 17–26 (2000).
69. Gupta, A., et al., Near infrared monitoring of roller compaction. *Proceedings — Institute for Briquetting and Agglomeration, Biennial Conference*, Santa Fe, New Mexico, **28**: pp. 65–78 (2004).
70. Gupta, A., et al., Real-time near-infrared monitoring of content uniformity, moisture content, compact density, tensile strength, and Young’s modulus of roller compacted powder blends. *J. Pharm. Sci.*, **94**: 1589–1597 (2005).
71. Gupta, A., et al., Influence of ambient moisture on the compaction behavior of microcrystalline cellulose powder undergoing uni-axial compression and roller-compaction: A comparative study using near-infrared spectroscopy. *J. Pharm. Sci.*, **94**: 2301–2313 (2005).
72. Miller, R.W., A. Gupta, and K.R. Morris, Roller compaction scale-up. *Drugs. Pharm. Sci.*, **157**(Pharmaceutical Process Scale-Up (2nd ed.)): 237–266 (2006).
73. Kirsch, J.D. and J.K. Drennen, Determination of film-coated tablet parameters by near-infrared spectroscopy. *J. Pharm. Biomed. Anal.*, **13**: 1273–1281 (1995).

74. Reich, G. and H. Frickel, Use of NIR transmission spectroscopy to determine physical and functional film coat properties on tablets. *Proceedings of the International Symposium on Controlled Release of Bioactive Materials*, San Diego, California, **26**: pp. 903–904 (1999).
75. Scheiwe, M.-W., D. Schilling, and P. Aebi, Near infrared spectroscopy analysis of intact pharmaceutical diclofenac coated tablets in transmission. *Pharm. Ind.*, **61**: 179–183 (1999).
76. Frickel, H. and G. Reich, NIR spectroscopy of film-coated tablets — fast and nondestructive evaluation of film coat uniformity and drug release kinetics. *Proceedings of the International Symposium on Controlled Release of Bioactive Materials*, Paris, France, **27**: pp. 740–741 (2000).
77. Dyrby, M., et al., Chemometric quantitation of the active substance (containing CN) in a pharmaceutical tablet using near-infrared (N-IR) transmittance and N-IR FT-Raman spectra. *Appl. Spectrosc.*, **56**: 579–585 (2002).
78. Wu, H., et al., Application of principal component analysis in assessing pharmaceutical formulation design: Exploring the casual links between the tablet processing conditions and drug dissolution rate. *AIChE Annual Meeting, Conference Proceedings*, San Francisco, CA, United States, November 16–21, pp. 385–392 (2003).
79. Andersson, M., et al., Quantitative analysis of film coating in a fluidized bed process by in-line NIR spectrometry and multivariate batch calibration. *Anal. Chem.*, **72**: 2099–2108 (2000).
80. Perez-Ramos Jose, D., et al., Quantitative analysis of film coating in a pan coater based on in-line sensor measurements. *AAPS PharmSciTech* [electronic resource], **6**: E127–E136 (2005).
81. Lodder, R.A., M. Selby, and G.M. Hieftje, Detection of capsule tampering by near-infrared reflectance analysis. *Anal. Chem.*, **59**: 1921–1930 (1987).
82. Lodder, R.A. and G.M. Hieftje, Analysis of intact tablets by near-infrared reflectance spectrometry. *Appl. Spectrosc.*, **42**: 556–558 (1988).
83. Jensen, R., et al., One-step quantification of active ingredient in pharmaceutical tablets using near-infrared spectroscopy. *Spectrosc. Int. J.*, **6**: 63–72 (1988).
84. Drennen, J.K. and R.A. Lodder, Nondestructive near-infrared analysis of intact tablets for determination of degradation products. *J. Pharm. Sci.*, **79**: 622–627 (1990).
85. Murray, I. and I.A. Cowe, Macro- and micro-specimen analysis by near infrared diode array spectroscopy. *Advances in near infrared spectroscopy*, ed. E. Stark. Weinheim, Germany: VCH (1991).
86. Monfre, S.L. and F.A. DeThomas, Non-destructive pharmaceutical tablet assay by near-infrared spectroscopy. *Near infrared spectroscopy: Bridging the gap between data analysis and NIR applications*, ed. K.I. Hildrum. Ellis-Horwood: Chichester, UK (1992).
87. Gottfries, J., et al., Vibrational spectrometry for the assessment of active substance in metoprolol tablets: A comparison between transmission and diffuse reflectance near-infrared spectrometry. *J. Pharm. Biomed. Anal.*, **14**: 1495–1503 (1996).
88. Merckle, P. and K.A. Kovar, Assay of effervescent tablets by near-infrared spectroscopy in transmittance and reflectance mode: Acetylsalicylic acid in mono and combination formulations. *J. Pharm. Biomed. Anal.*, **17**: 365–374 (1998).
89. Thosar, S.S., et al., A comparison of reflectance and transmittance near-infrared spectroscopic techniques in determining drug content in intact tablets. *Pharm. Dev. Technol.*, **6**: 19–29 (2001).
90. Cogdill, R.P., et al., Process analytical technology case study part I: Feasibility studies for quantitative near-infrared method development. *AAPS PharmSciTech* [electronic resource], **6**: E262–E272 (2005).
91. Trafford, A.D., et al., A rapid quantitative assay of intact paracetamol tablets by reflectance near-infrared spectroscopy. *Analyst* (Cambridge, United Kingdom), **124**: 163–167 (1999).
92. Guo, J.-H., et al., Application of near-infrared spectroscopy in the pharmaceutical solid dosage form. *Drug Dev. Ind. Pharm.*, **25**: 1267–1270 (1999).
93. Blanco, M., et al., Identification and quantitation assays for intact tablets of two related pharmaceutical preparations by reflectance near-infrared spectroscopy: Validation of the procedure. *J. Pharm. Biomed. Anal.*, **22**: 139–148 (2000).
94. Broad, N.W., et al., Application of transmission near-infrared spectroscopy to uniformity of content testing of intact steroid tablets. *Analyst* (Cambridge, United Kingdom), **126**: 2207–2211 (2001).
95. Ramirez, J.L., M.K. Bellamy, and R.J. Romanac, A novel method for analyzing thick tablets by near infrared spectroscopy. *AAPS PharmSciTech* [electronic resource], **2**: E11 (2001).
96. Mark, H., et al., Validation of a near-infrared transmission spectroscopic procedure, part A: Validation protocols. *J. Pharm. Biomed. Anal.*, **28**: 251–260 (2002).

97. Ritchie, G.E., et al., Validation of a near-infrared transmission spectroscopic procedure, part B: Application to alternate content uniformity and release assay methods for pharmaceutical solid dosage forms. *J. Pharm. Biomed. Anal.*, **29**: 159–171 (2002).
98. Laasonen, M., et al., Development and validation of a near-infrared method for the quantitation of caffeine in intact single tablets. *Anal. Chem.*, **75**: 754–760 (2003).
99. Smith, M.R., et al., Optimisation of partial least squares regression calibration models in near-infrared spectroscopy: A novel algorithm for wavelength selection. *Analyst* (Cambridge, United Kingdom), **128**: 1312–1319 (2003).
100. Tatavarti Aditya, S., et al., Assessment of NIR spectroscopy for nondestructive analysis of physical and chemical attributes of sulfamethazine bolus dosage forms. *AAPS PharmSciTech* [electronic resource], **6**: E91–E99 (2005).
101. Chalus, P., et al., Near-infrared determination of active substance content in intact low-dosage tablets. *Talanta*, **66**: 1294–1302 (2005).
102. Dou, Y., et al., Simultaneous non-destructive determination of two components of combined paracetamol and amantadine hydrochloride in tablets and powder by NIR spectroscopy and artificial neural networks. *J. Pharm. Biomed. Anal.*, **37**: 543–549 (2005).
103. Dou, Y., et al., Nondestructive quantitative analysis of cimetidine tablets using artificial neural networks in near-infrared spectroscopy. *Spectrosc. Lett.*, **38**: 1–11 (2005).
104. Dou, Y., et al., Determination of compound aminopyrine phenacetin tablets by using artificial neural networks combined with principal components analysis. *Anal. Biochem.*, **351**: 174–180 (2006).
105. Feng, Y.-C. and C.-Q. Hu, Construction of universal quantitative models for determination of roxithromycin and erythromycin ethylsuccinate in tablets from different manufacturers using near infrared reflectance spectroscopy. *J. Pharm. Biomed. Anal.*, **41**: 373–384 (2006).
106. Bodson, C., et al., Comparison of FT-NIR transmission and UV-vis spectrophotometry to follow the mixing kinetics and to assay low-dose tablets containing riboflavin. *J. Pharm. Biomed. Anal.*, **41**: 783–790 (2006).
107. Meza, C.P., A. Santos Maria, and J. Romanach Rodolfo, Quantitation of drug content in a low dosage formulation by transmission near infrared spectroscopy. *AAPS PharmSciTech* [electronic resource], **7**: E29 (2006).
108. Blanco, M. and M. Alcalá, Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy. *Anal. Chim. Acta*, **557**: 353–359 (2006).
109. Ebube, N.K., et al., Application of near-infrared spectroscopy for nondestructive analysis of Avicel powders and tablets. *Pharm. Dev. Technol.*, **4**: 19–26 (1999).
110. Gustafsson, C., et al., Characteristics of hydroxypropyl methylcellulose influencing compactibility and prediction of particle and tablet properties by infrared spectroscopy. *J. Pharm. Sci.*, **92**: 494–504 (2003).
111. Moffat, A.C., et al., Meeting the International Conference on Harmonisation's Guidelines on Validation of Analytical Procedures: Quantification as exemplified by a near-infrared reflectance assay of paracetamol in intact tablets. *The Analyst*, **125**: 1341–1351 (2000).
112. Harmonisation, I.C.o., ICH harmonised tripartite guideline-text on validation of analytical procedures. *Federal Register*, **60**: 11260 (1995).
113. Harmonisation, I.C.o., ICH harmonised tripartite guideline-validation of analytical procedures: Methodology. *Federal Register*, **60**: 27463 (1995).
114. Lodder, R.A. and G.M. Hieftje, Detection of subpopulations in near-infrared reflectance analysis. *Appl. Spectrosc.*, **42**: 1500–1512 (1988).
115. Zannikos, P.N., et al., Spectrophotometric prediction of the dissolution rate of carbamazepine tablets. *Pharm. Res.*, **8**: 974–978 (1991).
116. Drennen, J.K. and R.A. Lodder, Qualitative analysis using near-infrared spectroscopy: A comparison of discriminant methods in dissolution testing. *Spectroscopy*, **6**: 34–39 (1991).
117. Donoso, M. and E.S. Ghaly, Prediction of drug dissolution from tablets using near-infrared diffuse reflectance spectroscopy as a nondestructive method. *Pharm. Dev. Technol.*, **9**: 247–263 (2004).
118. Freitas, M.P., et al., Prediction of drug dissolution profiles from tablets using NIR diffuse reflectance spectroscopy: A rapid and nondestructive method. *J. Pharm. Biomed. Anal.*, **39**: 17–21 (2005).
119. Donoso, M. and E.S. Ghaly, Prediction of tablets disintegration times using near-infrared diffuse reflectance spectroscopy as a nondestructive method. *Pharm. Dev. Technol.*, **10**: 211–217 (2005).

120. Drennen, J.K. and R.A. Lodder, Pharmaceutical applications of near-infrared spectrometry. *Advances in near-infrared measurements*, ed. G. Patonay. Greenwich, CT: JAI Press (1993).
121. Morisseau, K.M. and C.T. Rhodes, Near-infrared spectroscopy as a nondestructive alternative to conventional tablet hardness testing. *Pharm. Res.*, **14**: 108–111 (1997).
122. Dempster, M.A., et al., Near-infrared methods for the identification of tablets in clinical trial supplies. *J. Pharm. Biomed. Anal.*, **11**: 1087–1092 (1993).
123. Aldridge, P.K., et al., Identification of tablet formulations inside blister packages by near-infrared spectroscopy. *Appl. Spectrosc.*, **48**: 1272–1276 (1994).
124. Lodder, R.A., M. Selby, and G.M. Hieftje, Near-infrared analysis of pharmaceuticals: Comparing quantile-beast and SIMCA for qualitative prediction. *Anal. Chem.*, **59**: 1921–1930 (1987).
125. Cogdill, R.P., A. Anderson Carl, and K. Drennen James, III, Process analytical technology case study, part III: Calibration monitoring and transfer. *AAPS PharmSciTech* [electronic resource], **6**: E284–E297 (2005).
126. Baxter, M., Aspects of Quantitative Analysis for Solid Dosage Forms. In *Eastern Analytical Symposium*. Somerset, NJ (1994).
127. Treado, P.J., I.W. Levin, and E.N. Lewis, Indium antimonide (InSb) focal plane array (FPA) detection for near-infrared imaging microscopy. *Appl. Spectrosc.*, **48**: 607 (1994).
128. Markwot, L., et al., Raman imaging of heterogeneous polymers: A comparison of global versus point illumination. *Appl. Spectrosc.*, **49**: 1411 (1995).
129. Clarke, F.C., S.V. Hammond, and C. Mattisson, The development of NIR microscopy for process control in pharmaceutical manufacturing. *J. Process Anal. Chem.*, **7**: 115–118 (2002).
130. Clarke, F.C., et al., Chemical image fusion. The synergy of FT-NIR and Raman mapping microscopy to enable a more complete visualization of pharmaceutical formulations. *Anal. Chem.*, **73**: 2213–2220 (2001).
131. Clarke, F., Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. *Vibrat. Spectrosc.*, **34**: 25–35 (2004).
132. Clarke, F.C., et al., Determination of the information depth and sample size for the analysis of pharmaceutical materials using reflectance near-infrared microscopy. *Appl. Spectrosc.*, **56**: 1475–1483 (2002).
133. Lyon, R.C., et al., Near-infrared spectral imaging for quality assurance of pharmaceutical products: Analysis of tablets to assess powder blend homogeneity. *AAPS PharmSciTech* [electronic resource], **3**: E17 (2002).
134. Medendorp, J. and R.A. Lodder, Applications of integrated sensing and processing in spectroscopic imaging and sensing. *J. Chemometr.*, **19**: 533–542 (2005).
135. Liu, J., Physical characterization of pharmaceutical formulations in frozen and freeze-dried solid states: Techniques and applications in freeze-drying development. *Pharm. Dev. Technol.*, **11**: 3–28 (2006).
136. Whitfield, R.G., Near-infrared reflectance analysis of pharmaceutical products. *Pharm. Manuf.*, **3**: 31–40 (1986).

31 NIR in the Dietary Supplement Industry: Qualitative and Quantitative Analysis of Ingredients, Process Blends, and Final Products

Cynthia Kradjel

CONTENTS

31.1 Challenges of Dietary Supplement Testing	614
31.2 MID-IR vs. NIR	616
31.3 Development of NIR Calibrations for the Identification and Quality Confirmation of Dietary Supplements	617
31.4 Typical Installation	621
31.5 Differentiation of Raw Root Powder vs. Extract	623
31.6 Quantitative Analysis	624
31.7 Chondroitin Sulfate	627
31.8 Detection of Adulteration	627
31.9 Literature	627
31.10 Future Developments	628
31.11 Conclusion	628
References	629

The development of near-infrared (NIR) for routine quality control (QC) testing in agricultural products by Karl Norris and coworkers in 1968 [1,2] and John Shenk [3,4] paved the way for successful NIR method development for nutraceutical applications. Natural products are complex substances that can vary widely, depending upon growing conditions, seasons, country of origin, and preprocessing. The current successful technique for dietary supplements combines advanced instrument hardware with sophisticated software using chemometric mathematical algorithms and calibration techniques.

Pioneering work in the 1980s by Emil Ciurczak [5] and Howard Mark [6] resulted in the acceptance of NIR technology in the pharmaceutical industry, primarily for identification and quality

confirmation, again paving the way for successful use in the dietary supplement industry with designs for regulatory compliance and with mathematical algorithms to effectively separate similar materials.

These developments come at a time when dietary supplement industry faces major changes in regulations with proposed Good Manufacturing Practices (GMP) unique to their business. On March 13, 2003, the Food and Drug Agency (FDA) published an article “Current Good Manufacturing Practice in Manufacturing, Packing, and Holding Dietary Ingredients and Dietary Supplements; Proposed Rule” [7]. The proposed rule is for anyone engaged in activities related to manufacturing, packing, or holding dietary ingredients or dietary supplements. The provisions would require manufacturers to evaluate the identity, purity, quality, strength, and composition of their dietary ingredients and dietary supplements.

American consumers are increasingly supplementing or replacing prescription medicines with medicinal herbs, vitamins, and minerals. An estimated \$20 billion was spent by U.S. Consumers on natural supplements in 2004 [8]. In addition, the consumption of functional foods, nutraceutical drinks, and other types of dietary supplements are growing at a rapid pace. Manufacturers as well as marketing companies and suppliers require implementation of analytical techniques for these products. NIR technology is increasingly being adopted as the testing method of choice. NIR is fast, easy to use, nondestructive, and requires almost no sample preparation. Both physical and chemical properties can be measured. Testing can take place in the lab or warehouse. Once a method has been developed, routine testing takes place in a few seconds, making it possible to test every drum of incoming ingredient to verify the identity and quality.

31.1 CHALLENGES OF DIETARY SUPPLEMENT TESTING

The application of NIR technology to the dietary supplement industry is indeed analytically challenging for many reasons:

1. The physical state of incoming raw materials spans the range from clear liquids to opaque liquids or viscous materials, powders to tablets or capsules, syrups, or whole products. Sample handling devices have been developed to facilitate this range of physical forms. System configurations include fiber optic sampling or window-based sampling. Fiber optics enables users to probe in the drum directly, through a plastic sleeve or through plastic drum liners to eliminate cross-contamination concerns and to eliminate the need to open packaging. Noncontact testing techniques prevent the risk of contamination and worker exposure. With rapid testing, customers can perform identify and purity testing while tracking every drum of raw materials received right in the warehouse area. On the other hand, window-based sampling enables companies to implement a system that includes taking retained samples from drums, storing them in glass vials, and then placing the glass vial on the optical window to obtain the spectrum.
2. Product proliferation — it is common for dietary supplement manufacturers to have over 1000 different types of raw materials in their warehouses, particularly contract manufacturers who process for a variety of customers. Products tend to move quickly into and out of the marketplace. Method development must be flexible, rapid, and responsive. Successful NIR users in the dietary supplement industry have programs in place to routinely update their databases and calibrations.
3. Natural products have complex matrices that may contribute to their benefit for the consumer. However, as the materials are being analyzed “as is” without any physical separation of the analyte from the matrix, identity tolerances developed on a material with a specific binder will fail a new material if a substitute binder has been used. Calibration procedures and Standard Operating Procedures (SOPs) must be developed to ensure that

- (a) similar materials can be separated and are not misidentified, and (b) procedures are in place to diagnose a material that “fails.” Is the material the wrong substance or has the supplier changed some aspect of the processing? A material that falls outside of the acceptable tolerance level is either sent back to the manufacturer or analyzed by reference methods to find out what the differences really are between the new batch and the previously acceptable batches. Sometimes, the material is determined by primary methods to be of acceptable quality and then the criteria for passing material are updated in the calibration to include the profile of the new material.
4. Natural products vary from batch to batch. Mathematical algorithms such as cluster analysis can be used to set boundaries that allow acceptable variation within a product class. Each product class will have different profiles with the most homogeneous and reproducible material having small tolerance level on the cluster, and with materials with wide acceptable variation having larger tolerance level on the cluster.
 5. It is critical to test differences between different parts of the plant or differences between original material and extracts. Physical and chemical differences existing in roots, stems, or leaves can be measured by NIR. For manufacturers preparing extracts, it is useful to monitor the target points in their production with NIR. For customers purchasing extracts, it is important to verify if they have paid for extracts, and that they are indeed receiving extracts rather than a modified or adulterated material that appears to be an extract.
 6. Many companies have been sending materials out for testing and have not built in-house laboratories. For smaller companies facing compliance to the pending regulation, the prospect of building and staffing a full laboratory is rather daunting. On the other hand, relying solely on outside testing labs can mean less overall testing capability, time delays, and added expense. NIR equipment can be set up on a cart or mobile workstation, and easily operated in routine use. As it is a nondestructive technique, there is no need for hazardous solvents to perform testing. Lot-to-lot consistency can be ensured and sub-par and adulterated materials can be detected. All incoming containers can be quickly tested to assure that they are the correct material and are within user-defined specifications.
 7. Purity testing is becoming more important as demand for botanicals, herbals, and natural products grow and supply issues put pressure on raw material suppliers. Manufacturers must look for the presence of spiked or doped material. Similar raw materials can often fool traditional tests, so manufacturers order a \$2000/kg raw material and may receive a \$200/kg similar raw material that may be difficult to tell apart using traditional laboratory methods. Full spectral testing can also detect ingredients that have been adulterated with undisclosed and unwanted materials such as synthetics.
 8. Companies should implement a vendor certification program. Cluster analysis successfully provides unambiguous results for complex botanical materials, using factor analysis that represents the complete spectrum of the material as a point in factor space. Operators are not expected to interpret the spectral patterns as the software performs the calculations and reports the result. It is now possible to obtain Fourier transform near infrared (FT-NIR) fingerprints of all incoming and outgoing materials. “More than 70% of manufacturers switched raw ingredients suppliers last year — 51.4% cited quality as why” (Nutritional Outlook Study, April 2002) [9]. The manufacturer is responsible for ensuring that the supplements are safe and effective, that they possess the appropriate identity, strength, and purity as they are represented on the final product label.
 9. Documentation of quality of incoming ingredients, as well as outgoing ingredients, is a means to protect oneself against frivolous lawsuits or unwarranted product claims.
 10. Most manufacturers recognize the need to be GMP certified, to streamline raw material handling, and to prepare for the pending FDA regulations. The FT-NIR solution is easy to implement, simple to use, and powerful enough to keep up with changing regulatory requirements. Companies using the system know that reliability, accuracy, and compliance

can be achieved quickly and easily. With roots in the pharmaceutical industry, systems are on the markets that comply with 21 CFR 11 guidelines, the food and pharma GMP regulations, as well as the proposed dietary supplement regulation. In the USP Chapter 1119, NIR has been recognized as an acceptable method when equivalency is demonstrated to the primary methods.

Americans must have confidence that the dietary supplements that they purchase are not contaminated and that they contain the dietary ingredients and the amounts claimed on the labels. Millions of Americans use Dietary Supplements and we owe it to them to ensure that they are getting the products that they're paying for [10].

Tommy G. Thompson, Secretary, Department of Health and Human Services

The challenges and requirements for protecting a company's health are numerous. Governmental regulations now being implemented are requiring a dramatic increase in the testing of dietary supplements [11].

Food Processing, "Get What You Pay for," September 2000

Suppliers and manufacturers as well as marketing and distribution companies are implementing FT-NIR to address the above-listed issues and to meet current and future regulatory requirements. The typical scenario for FT-NIR implementation in manufacturing is described in the article "Pharmavite: 2001 Manufacturer of the Year" (*Nutritional Outlook*, November/December 2001) [12].

When ingredients arrive at the warehouse, operators check the shipment against the purchase order. The computer logs the material into stock and generates a receiver and a bar code tag to identify the material through processing.

Next, all materials are weighed. The computer prints an outweigh verification report that confirms that the order is correct, and this becomes part of the permanent batch record. The first analytical check begins as soon as raw materials are received on the dock. To confirm the material is what it claims to be, a quality control (QC) technician performs nondestructive FTNIR spectroscopy. A special stainless steel probe is inserted into each sample. If the test fails, the material is quarantined and further analyzed in the lab. In addition, the QC technician pulls samples based on the square root of the number received minus one. For example, in a shipment of nine drums, the technician pulls four samples. After confirmation of identity, materials are put on hold and stored in the warehouse. When quality assurance is satisfied with the results of its test, the material is released for processing.

Suppliers, on the other hand, typically implement the system in their offices, and request pre-shipment samples from their suppliers to scan for compliance before approving the full shipment for release. They also use it to qualify new suppliers, ensuring that the physical and chemical properties of the new material are consistent with what they have been providing their customers. Before shipment of a raw material to their customers, they often scan again to have a record of what the spectral fingerprint of the material looked like before shipping.

Marketing and distribution companies often purchase raw materials from suppliers and then arrange them to be sent to a contract manufacturer for processing. These companies are implementing FT-NIR as a means to control the quality of the product that bears their label by checking not only on the ingredient quality, but also the output from the contract manufacturer.

31.2 MID-IR VS. NIR

The bands in the NIR region are overtones and combination bands of fundamental Mid-IR (MIR) spectra. There are several advantages of using NIR over MIR for this application. First, the Mid-IR measures mainly fundamental vibrations in the fingerprint region. Owing to high values of extinction,

the substances have to be “diluted” with KBr or Nujol. This method is time-consuming and not suitable for measuring a high number of samples in a short time. The NIR range exhibits overtones and combination vibrations with values of extinction less than 10% of the values in the Mid-IR. This fact allows the measurement of substances in a fast and direct method without sample preparation. NIR spectrophotometers also work well with fiber optics and are able to measure without the need for sample collection. The time for measuring one spectrum, with identification, is typically less than 10 s. In addition, FT-NIR light penetrates deeper into the sample. As there is no sample preparation, both chemical and physical properties are measured.

31.3 DEVELOPMENT OF NIR CALIBRATIONS FOR THE IDENTIFICATION AND QUALITY CONFIRMATION OF DIETARY SUPPLEMENTS

Numerous advantages of using NIR for testing of dietary supplements have been described. However, it is important to recognize that in order to obtain these results, methods and calibrations must be developed requiring access to samples that have been already tested by primary reference methods.

The simplest form of using spectroscopy is to simply scan materials and overlay the spectra. The usual approach for comparing a sample with a reference by eye is to hold their plotted spectra on top of one another in a kind of overlay technique. The identity is given according to the U.S. Pharmacopoeia if the absorption maxima are at the same wavelengths. This method is applicable to materials that are sufficiently different. However, this method would most likely be unable to discern small differences between materials or detect adulteration.

A second approach is to use software that has the ability to calculate how closely one spectrum correlates with the reference. This is more useful than visual comparison, but it is still subject to error and misinterpretation if the differences between the two materials do not result in distinct peak changes.

When FT-NIR data is combined with chemometrics, the increase in the amount of information available from the spectra are substantially increased, including the ability to separate very similar materials and to differentiate between materials of various grades and qualities. Chemometrics provides the algorithms for software to perform pattern recognition on the spectra. The most discerning method, Cluster Analysis, is based on software that uses factor analysis or principal components. This method is more objective as circles of tolerance are built with about five lots of typical material that has been tested according to reference methods.

When an unknown is scanned, the mathematical algorithms calculate if the unknown fits into any known circles of tolerance. The operator is not expected to interpret anything from the reading; a pass or fail result is provided from the software algorithm calculations. The entire spectral range is scanned, and during the calibration process, the software deconvolutes the spectra into components or factors.

The spectrum is deconvoluted according to the given components and the calculated loadings determine the point in the factor space. To compute the distances to all the references in the library, and searching for the most similar reference is quite a fast and easy task for a computer. Figure 31.1 shows a three-dimensional plot of the 5 different nutraceutical ingredients with about 25 spectra of each. The spectra are information rich. Figure 31.2 represents original spectra for multiple lots of agnus castus berry, dong quai, ginger, and turmeric. Although differences between the groups can be seen by eye in the original spectra, it is easier to discern the differences by eye when viewing spectra that have undergone a first-derivative transformation (Figure 31.3). Cluster analysis has been performed on this data, and Figure 31.4 shows the four different clusters for these materials in three-dimensional space, clearly showing that the groups can be separated. The cluster analysis calculations result in two mathematical measurements: Mahalanobis distance and residual.

The Mahalanobis distance tests whether or not an unknown fits within the limits of any known groups. The residual represents the spectral information that remains after the principal components

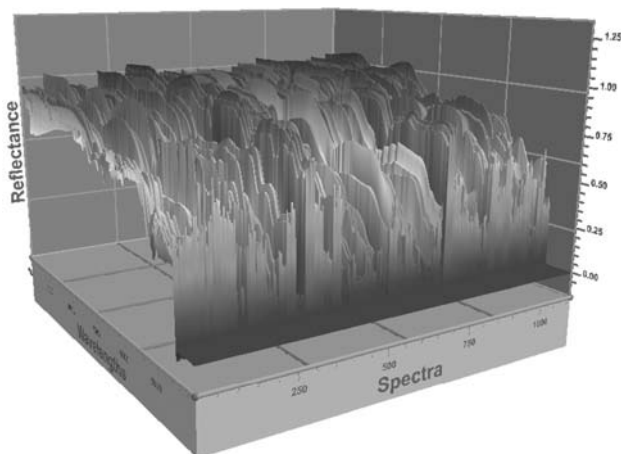


FIGURE 31.1 (See color insert following page 622.) Three-dimensional plot of 5 different nutraceutical ingredients with about 25 spectra of each.

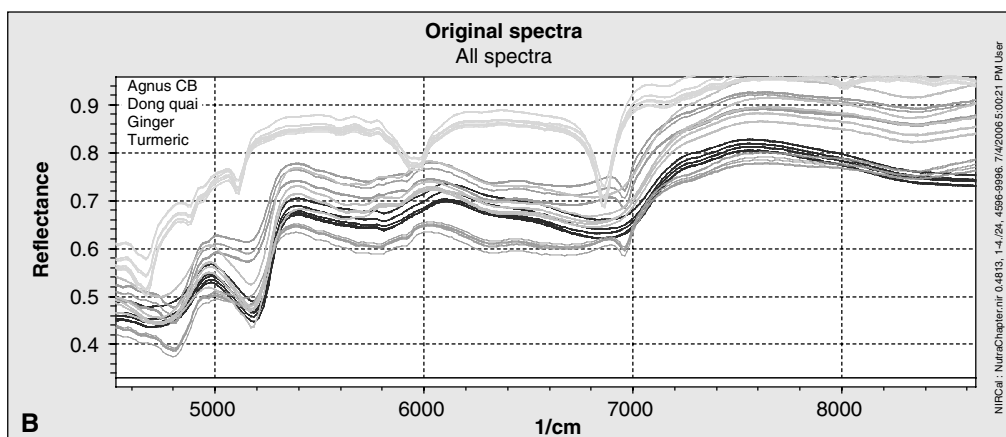


FIGURE 31.2 (See color insert following page 622.) Raw spectra for multiple lots of agnus castus berry, dong quai, ginger, and turmeric.

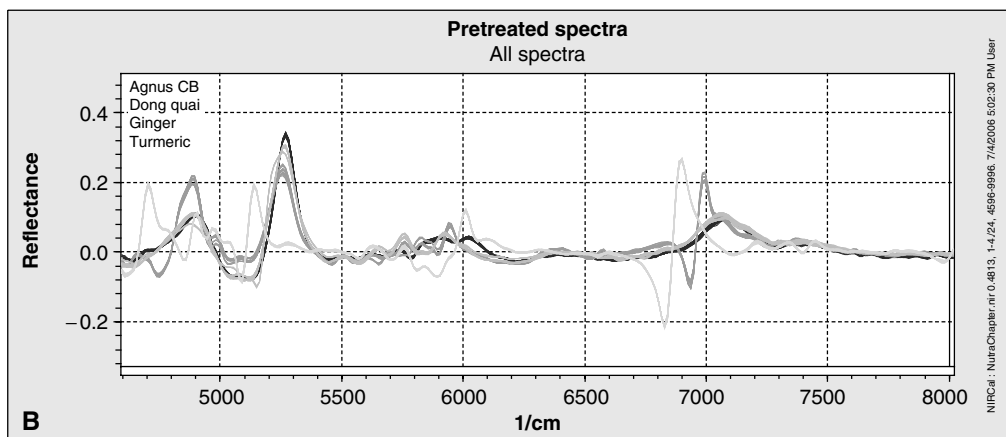


FIGURE 31.3 (See color insert following page 622.) First derivative spectra for multiple lots of agnus castus berry, dong quai, ginger, and turmeric.

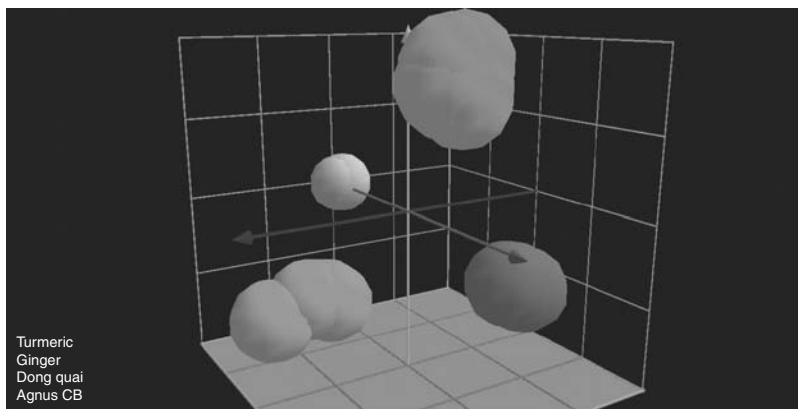


FIGURE 31.4 (See color insert following page 622.) Clusters for four different materials in three-dimensional space, clearly showing that the groups can be separated.

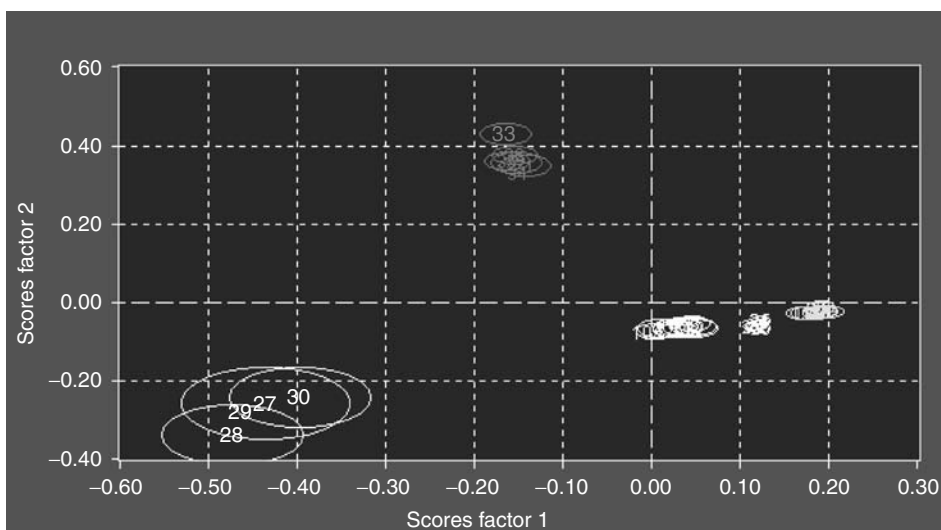


FIGURE 31.5 Figure shows various grades of vitamin C with the cluster of the unrefined vitamin C occupying the highest cluster.

have been subtracted from the spectra. Similar materials should have the same residual, and if a material has been adulterated or has unusual impurities, the residual spectra will look different than normal, and the system will fail the material. With this kind of data compression, the identification of an unknown sample against a library is fulfilled in only a few seconds. Substances are compared with their nearest neighbors; this enables separation of very similar materials and distinction of grade and quality. The cluster method is effective for separating various grades of quality. Figure 31.5 shows various grades of vitamin C, with the unrefined cluster occupying the greatest volume and the highly refined vitamin C occupying the smallest volume.

NIR spectra taken from substances in powder form without sample preparation are affected by grain size and packing density. This will cause slight variability in the spectra and therefore in the position of the correlated point in factor space. Clusters that have much higher variability in the position are the result of changing amounts of impurities from batch to batch. This variability is then tested by the measurement of different batches of the same substance. If the results show that the average batch-to-batch variability is significantly smaller than the average distance between other references, then the material is acceptable for testing in routine use.

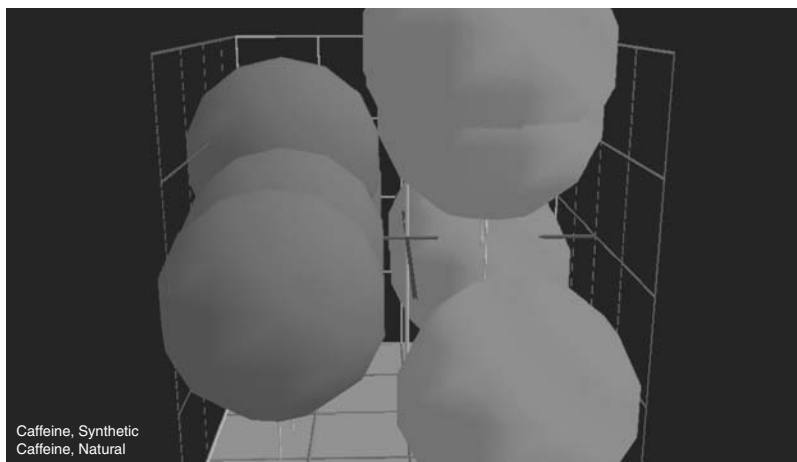


FIGURE 31.6 (See color insert following page 622.) Clusters of natural caffeine and synthetic caffeine.

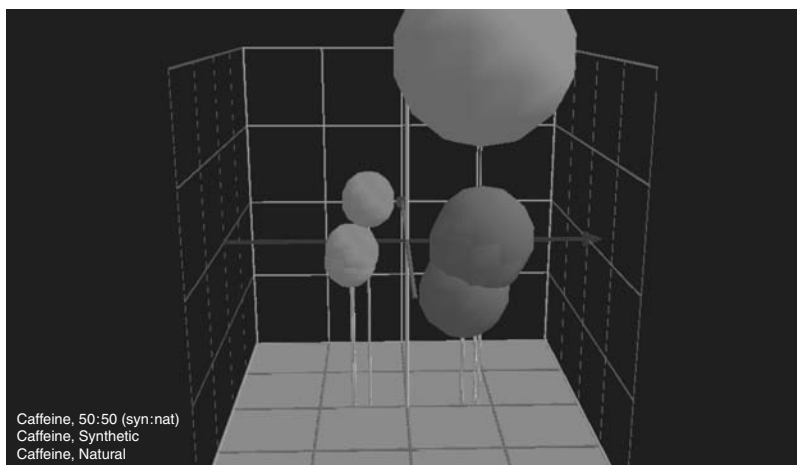


FIGURE 31.7 A 50:50 mixture of natural and synthetic caffeine does not fall into either the 100% natural caffeine cluster nor the 100% synthetic caffeine cluster.

Figure 31.6 shows a cluster analysis of synthetic and natural caffeine. Although the chemical entity is the same, the processing conditions are sufficiently different to allow separation between the two materials, based on teaching the system multiple samples of both products. Figure 31.7 shows that a 50:50 mix of synthetic and natural caffeine does indeed fall outside the range of both the natural and synthetic clusters.

Changing amounts of impurities causes a scatter in the points in factor space, each of which represent a spectrum. Some changes are tolerable; others are not. For example, a specific substance is required to have water content in a certain range. Substances with water content in this allowed range give points that lie close together and form a cluster in factor space, while in the other substances the points lie outside. If enough samples are available with known but not necessarily quantitatively determined properties (e.g., good and bad quality), and these properties are measurable in the range of NIR, a clustering for the allowed substances is possible. Manufacturers purchase raw materials according to agreed upon specifications, so it is important to check that the supplier is not providing material with water concentrations higher than those allowed on the specification; otherwise

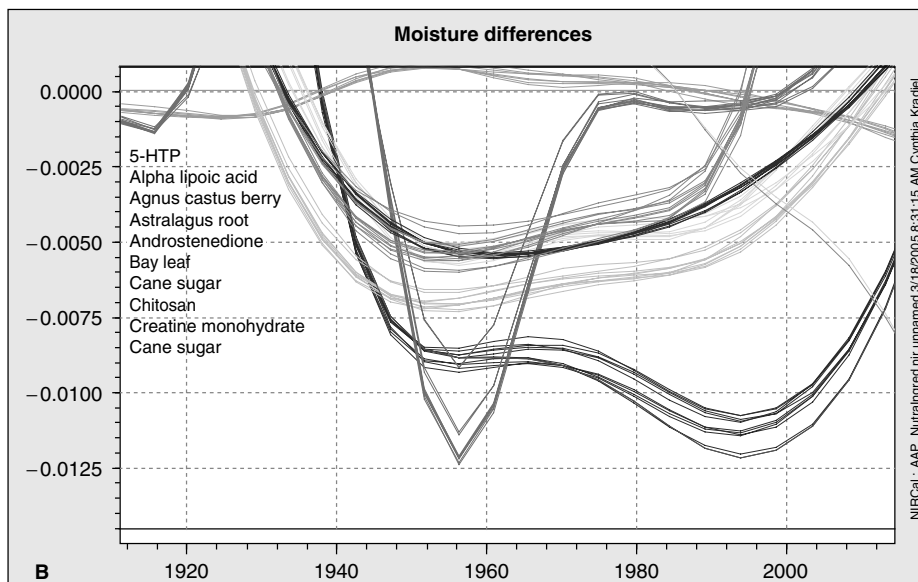


FIGURE 31.8 Various nutraceutical materials with varying moisture content, including some examples of materials with out-of-specification moisture content.

the manufacturer is in effect paying for water. Figure 31.8 shows various nutraceutical materials with varying moisture content, including some spectra with out-of-specification water content. Bound as well as free water can be differentiated with NIR.

Cluster analysis provides a tool for separation of very similar materials, and the success of these methods is highly dependent upon the calibration techniques used to build the model. Key elements of experimental design include the selection of samples used to teach the system and the performance of their corresponding reference tests. Proper grouping of the raw materials into appropriate cluster models is also important. To separate close materials, similar materials should be grouped together and then cross-validated against the remainder of materials in the database.

31.4 TYPICAL INSTALLATION

All American Pharmaceutical Company, in Billings, Montana, has successfully implemented FT-NIR for routine testing of identity and quality of all inbound raw materials as ingredients for health foods and nutritionally beneficial formulations [13]. Every container entering the facility is tested to guarantee that identity, purity, safety, and potency standards are met. The containers are sampled and then quarantined until completely approved. Once testing is completed and the material is approved, each ingredient is issued a Certificate of Analysis (COA), which corresponds to each lot number. Their products range from a wide variety of health food ingredients, herbals, and nutritional supplements. The products range in physical form as solids, liquids, gels, and tablets. Qualitative methods are used for identity and quality testing, and quantitative methods are used for in-process optimization and for generating data for COA.

All American Pharmaceutical Company analyzes more than 100 herbals with the FT-NIR. The herbals are grouped into several library models. Figure 31.9 shows the reference spectra for herbals in one of the herbal libraries. Mathematical data treatment such as first derivative makes the differences more noticeable to the eye. The mathematical algorithms view the data in multivariate space and are able to resolve very small differences that are not obvious to the human eye. Figure 31.10 shows the same materials in Figure 31.9 now graphed using first-derivative data treatment. Other

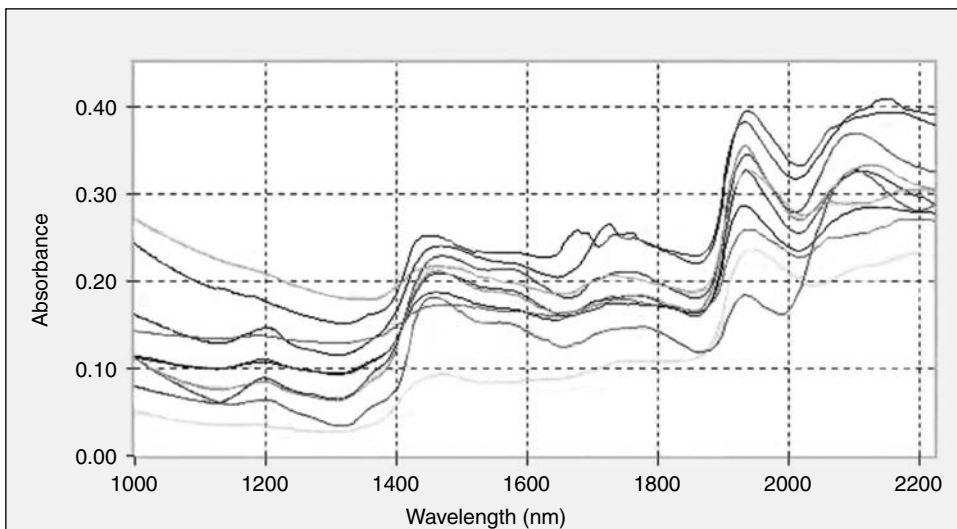


FIGURE 31.9 (See color insert following page 622.) Spectra of various herbal materials.

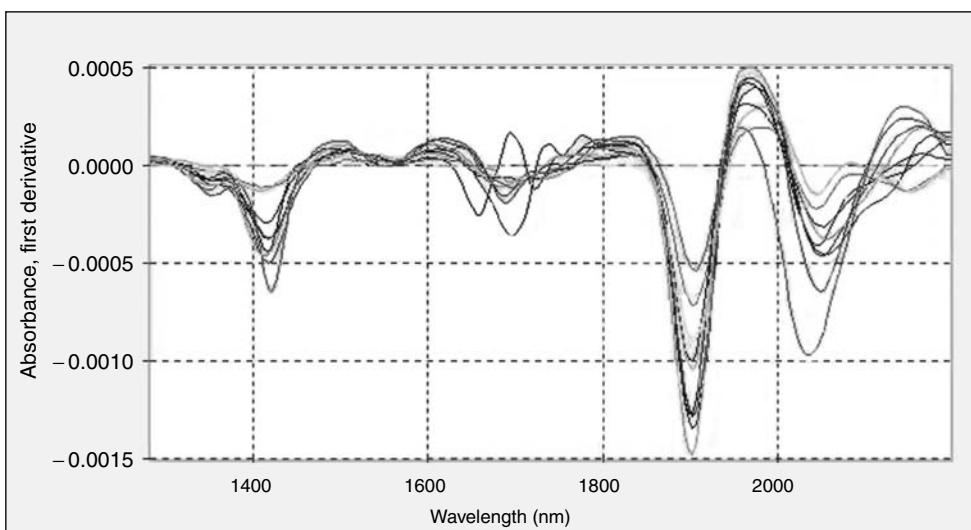


FIGURE 31.10 (See color insert following page 622.) First-derivative spectra of various herbal materials.

libraries in use at All American Pharmaceutical include amino acids (Figure 31.11), carbohydrates (Figure 31.12), vitamins (Figure 31.13), and sugars (Figure 31.14). Miscellaneous materials containing a variety of molecular species, which were easy to distinguish, were grouped into another model. This testing program assures proper ingredient identification and prevents formulation with incorrect or defective ingredients.

Rubinovitz [14] studied a variety of botanicals including chickweed, echinacea, golden seal, echinacea root, golden seal herb, golden seal root, horsetail, and St. John's Wort. Materials were scanned with a fiber-optic probe, and although the spectra of the powders were very similar when evaluated by eye, the cluster algorithm permitted reliable use of subtle differences in the powders' spectra to identify each type of powder and to accurately distinguish it from the other powders. He found that cluster methods covering larger numbers of similar products require more cluster-analysis weighting or loading factors in order to distinguish different materials from one another.

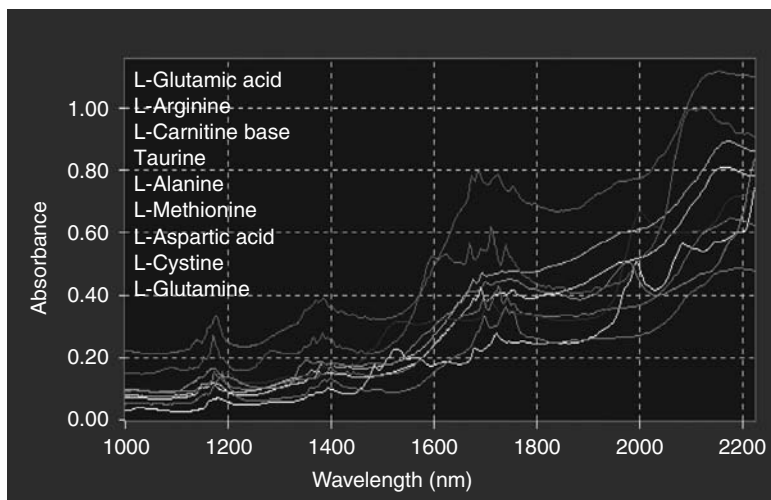


FIGURE 31.11 Original spectra of various amino acids.

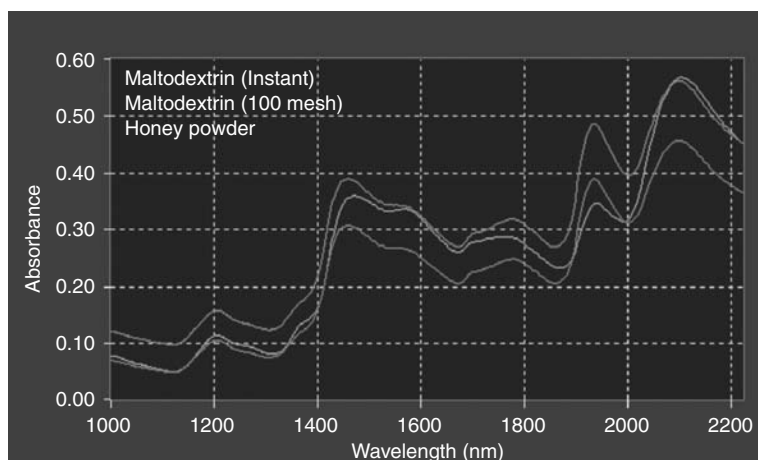


FIGURE 31.12 (See color insert following page 622.) Original spectra of various carbohydrates.

Three weighting factors were needed to separate the clusters for each of the original seven botanical powders described above.

31.5 DIFFERENTIATION OF RAW ROOT POWDER VS. EXTRACT

Another interesting example of using cluster analysis to separate materials that are spectrally similar is the FT-NIR differentiation of raw burdock root powder and (concentrated) burdock root extract [13]. An extraction process concentrates the active ingredients of a raw material. Dried extract (powder form) is often used to provide concentration amounts of active ingredients in final products. Figure 31.15 shows the difference between the spectra of the raw burdock root powder and the burdock root extract. Figure 31.16 shows a graphical representation of the different groups (raw burdock root and burdock root extract) known as a three-factor cluster plot. The variation in the burdock root is far greater than the variation in the burdock root extract as seen by the varying sizes of their

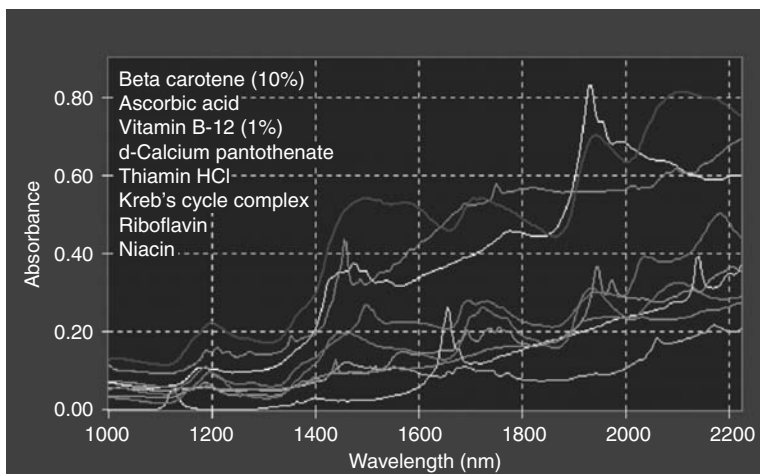


FIGURE 31.13 (See color insert following page 622.) Original spectra of various vitamins.

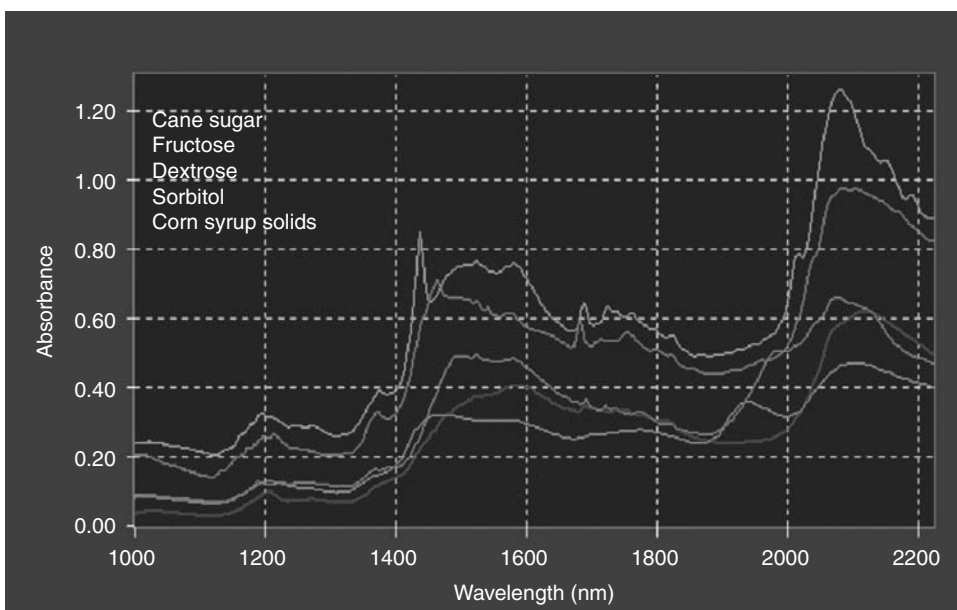


FIGURE 31.14 (See color insert following page 622.) Original spectra of various sugars.

respective clusters. This method can be used to quickly and accurately confirm that a material expected to be an extract is truly an extract. Cocoa Powder and Cocoa Extract have been studied, and can be differentiated. Figure 31.17 shows cocoa powder and cocoa extract. Additional information can be seen from the spectra, including the unexpected presence of an organic solvent, methanol, used in the extraction process (Figure 31.18).

31.6 QUANTITATIVE ANALYSIS

Quantitative methods have also been developed at All American Pharmaceutical [14]. For example, the assay of 19-Norandrosterone was built by using a population of about 20 teaching samples with

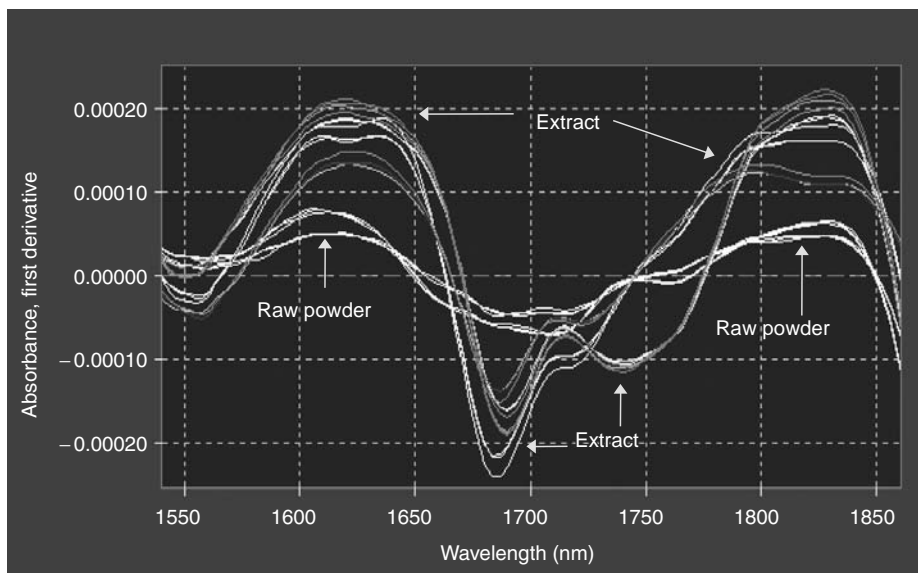


FIGURE 31.15 (See color insert following page 622.) First-derivative spectra of raw burdock root and burdock root extract.

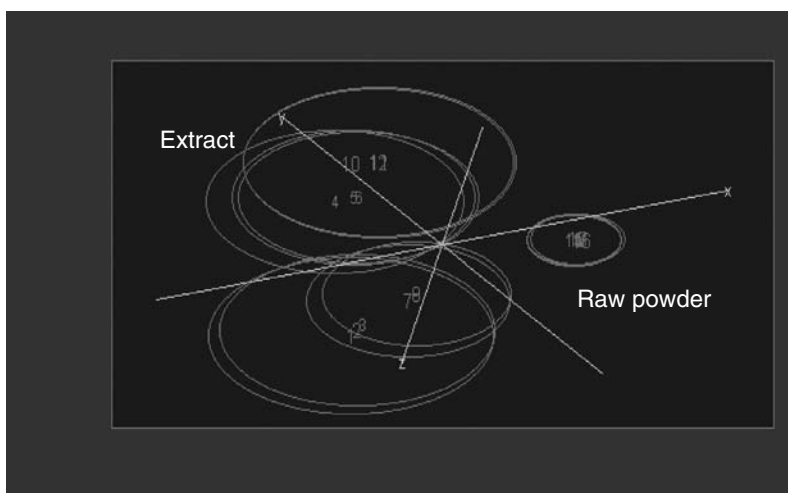


FIGURE 31.16 Cluster separation of raw burdock powder and burdock extract.

varying references. Partial Least Squares, Multiple Linear Regression, and Principal Component Analysis mathematical algorithms are used to evaluate the data and generate a correlation of the spectra to the reference method. This creates a formula that can be used to determine the assay of unknown materials of the same product type as those used to teach the system. Figure 31.19 shows spectra of varying concentration of Assay of 19-Norandrosteine. The reference method was HPLC. The FT-NIR was able to predict unknown 19-Norandrosteine Assay in the range of 92 to 102% assay (Figure 31.20).

Gray et al. [15] developed an NIR method to quantify chicoric acid in purple coneflower roots. One hundred and sixty-nine plants were harvested and their root tissues were scanned in NIR and also analyzed for chicoric acid by HPLC. A correlation between the HPLC spectra and the NIR data was 0.9, with a mean chicoric acid concentration of 8.29 g/kg DM and standard errors of calibration of 0.89.

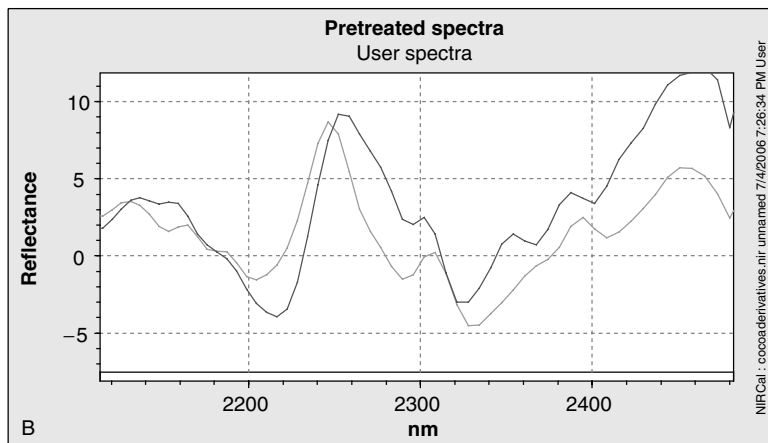


FIGURE 31.17 Cocoa powder vs. cocoa extract.

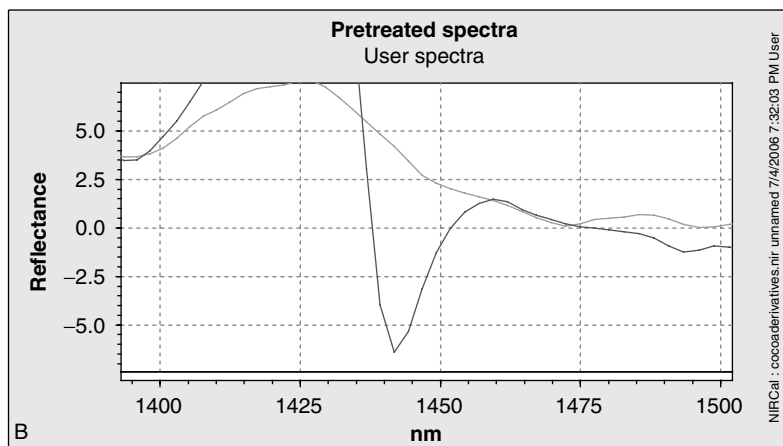


FIGURE 31.18 Cocoa extracts, one of which contains an unexpected residual solvent, methanol.

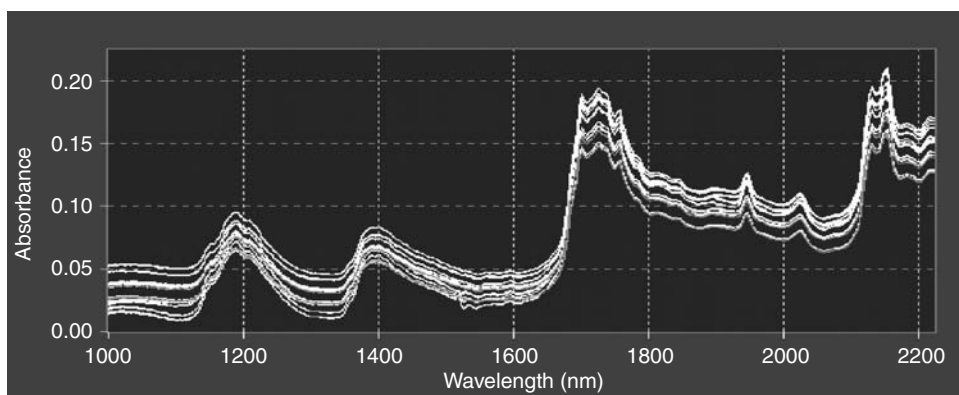


FIGURE 31.19 (See color insert following page 622.) Original spectra of 19-Norandrosterone Assay.

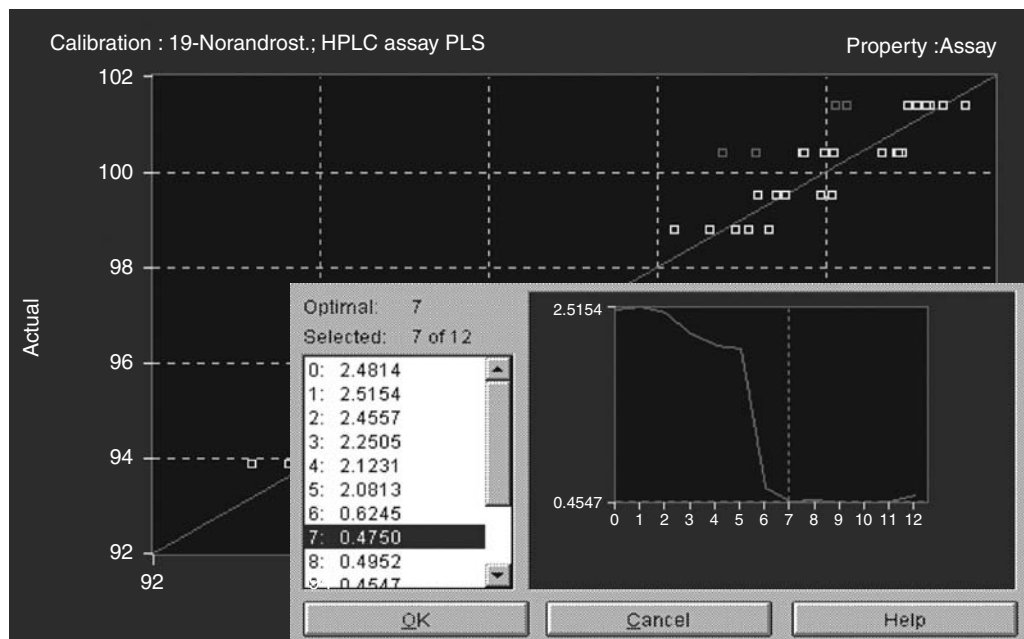


FIGURE 31.20 Quantitative method for 19-Norandrosterone Assay.

31.7 CHONDROITIN SULFATE

Yatcilla [11] found that chondroitin sulfate exemplified the danger of relying only on marker compounds. For chondroitin sulfate, a sulfated polysaccharide, the marker compound is notoriously difficult to assay by traditional means. Analytical testing of chondroitin sulfate has a history of false-positive readings. Figure 31.21 shows three different lots of chondroitin sulfate with corresponding results from HPLC, CE, FT-NIR, and the Carbazole method. Although all three lots passed the HPLC and the CE tests, two of the three failed the FT-NIR method. The Carbazole test confirmed that the results obtained on the FT-NIR were accurate. False-positive HPLC results typically indicate that another material is co-eluting at the same time as the material being analyzed.

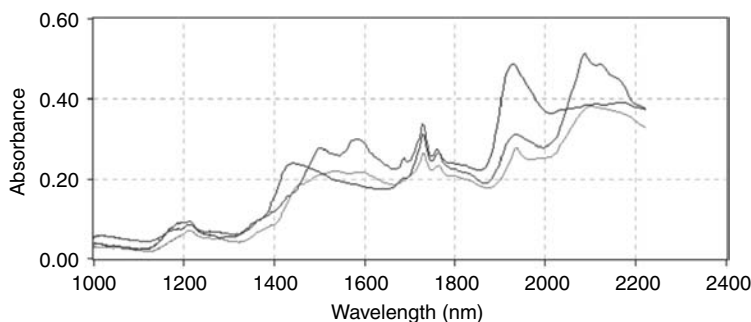
31.8 DETECTION OF ADULTERATION

Juliani et al. [16] used NIR to determine adulteration of essential oils. An adulterated sample composed of 1,8-cineole with authentic ravsensara (*Rhus aromatica*) oil failed the test for identity, demonstrating that this technique has the ability to identify adulterated essential oils. They also demonstrated that NIR could distinguish essential oils of different chemotypes and species.

Golini [17] studied the use of NIR for detection of adulteration of botanicals with synthetic drugs, including Ma Huang containing synthetic ephedrine HCl, kola nut, and guarana nut containing synthetic caffeine anhydrous, and *Sida cordifolia* containing synthetic ephedrine HCl. Figure 31.16 shows a cluster model of 100 drums tested, of which some passed as 100% natural and some in the shipment were adulterated with synthetic ephedrine HCl.

31.9 LITERATURE

Additional work reported in the literature includes a study on columbianadin by Abraham et al. [18], the discrimination of herbal medicines according to geographical origin with NIR and pattern



Sample	HPLC	CE	FT-NIR	Carbazole
Red	91.5%	101.6%	Fail	14.6%
Green	96.1%	N/A	Fail	11.4%
Blue	95.8%	99.7%	Pass	99.7%

FIGURE 31.21 (See color insert following page 622.) FT-NIR detects materials that give false positive by other techniques for chondroitin sulfate.

recognition by Young-Ah Woo et al. [19,20], and discussion of QC in the manufacture of herbal medicinal products by Laasonen [21]. Schulz et al. [22] performed rapid classification of basil chemotypes by various vibration spectroscopy methods, and Laasonen et al. [23] reported “Fast Identification of *Echinacea purpurea* Dried Roots Using Near-Infrared Spectroscopy.” Kudo et al. [24,25] reported using NIR for identification including *Digitalis purpurea*. Pudelko-Koerner [26] reported quantitative analysis of sennosides from *Sennae fructus angustifoliae* for in-process and quality-control testing.

31.10 FUTURE DEVELOPMENTS

As mentioned above, the FT-NIR cluster method for testing incoming ingredients does not depend on purchasing “standards”; rather it is achieved by using about five lots of material that represent the quality of material considered acceptable. These five lots are determined to be acceptable by performing USP or reference tests and then determining equivalency of the FT-NIR method to these tests. There has been considerable discussion particularly with botanicals about methods to ensure that the appropriate marker compounds are present in an ingredient, and, therefore, work is underway to collaborate with experts in botanical identification at ChromaDex in California to scan materials in the NIR that have been certified to contain the appropriate level and type of marker compounds.

The sensitivity of NIR discrimination is a function of the cluster method parameters used, and NIR has demonstrated that it can successfully monitor small differences in materials by the chemometric analysis of the NIR patterns. Databases of materials that have been properly classified with reference methods are critical to the growth of NIR for use in botanicals.

31.11 CONCLUSION

NIR technology can successfully be used to measure identity, quality, and quantity aspects of dietary supplements. Chemometrics is a proven technique for interpreting the spectral patterns and providing unambiguous results for routine operation. Results are rapid and nondestructive, enabling manufacturers to rapidly and nondestructively test incoming raw materials for adulteration, for identity, and quality confirmation. Manufacturers can now test not only pre-lot shipments, but also each individual drum upon receipt, verifying lot-to-lot and batch-to-batch consistency.

NIR methods can be adjusted as necessary to adapt to the changing needs of this industry by providing rapid identity and quality confirmation. NIR can also verify the quality and grade of incoming raw materials, to ensure product consistency and to facilitate compliance with regulatory guidelines including 21 CFR 11 and cGMPs. Quantitative analysis of active ingredients and other important properties of raw materials, in-process materials, and final products can be determined effectively. The industry overall will realize costs savings in sourcing and production while leveling the playing field for quality issues.

REFERENCES

1. Ben Gera, I. and Norris, K.H., *Isr. J. Agr. Res.*, **18**, 125 (1968).
2. Ben Gera, I. and Norris, K.H., *Journal of Food Science*, **33**, 64 (1968).
3. Shenk, J.S. and Hoover, M.R., "Infrared Reflectance Spectro-Computer Design and Application," in *Proc. 7th Technicon Int. Cong.*, 2, Tarrytown, NY, 1976, p. 122.
4. Shenk, J.S., Barnes, R.F., and Moore, J.E., *Journal of Animal Science*, **43**, 889 (1976).
5. Ciurczak, E.W., *Proc. 7th Annual Symposium on NIRA*, Technicon, Tarrytown, New York, 1984.
6. Mark, H.L. and Tunnell, D., *Analytical Chemistry*, **57**, 1449 (1985).
7. "Current Good Manufacturing Practice in Manufacturing, Packing and Holding Dietary Ingredients and Dietary Supplements; Proposed Rule," *Federal Register*, Department of Health and Human Services, Food and Drug Administration, 21 CFR Parts 111 and 112, March 13, 2003.
8. Nutrition Business Journal, NJB's Supplement Business Report 2005.
9. Nutritional Outlook Study, April 2002.
10. Department of Health and Human Services Secretary, Tommy G. Thompson.
11. Best, D., "Get What You Pay For," *Food Processing*, 82–84, September (2000).
12. Wagner, J., "Pharmavite: 2001 Manufacturer of the Year," *Nutritional Outlook*, November/December (2001).
13. Kradjel, C., Muller, D., Eilert, A., and Golini, J., "Qualitative and Quantitative NIR Spectroscopic Evaluation of the Raw Ingredients for Health Foods and Nutritionals," Pittcon, Orlando, Florida, March 1999.
14. Rubinovitz, R., "Rapid Identification of Raw Materials with Near-Infrared Spectroscopy," *Natural Pharmacy*, **6**(4), April (2002).
15. Gray, D.E., Roberts, C.A., Rottinghaus, G.E., Garret, H.E., and Pallardy, S.G., "Quantification of Root Chicoric Acid in Purple Coneflower by NIR Spectroscopy," *Crop Science*, **41**, 1159–1161 (2001).
16. Juliani, H.R., Kapteyn, J., Jones, D. et al., "Application of Near-Infrared Spectroscopy in Quality Control and Determination of Adulteration of African Essential Oils," *Phytochemical Analysis*, **17**, 121–128 (2006).
17. Company Tests Herbs for Drug Content, *World Foods Magazine*, April 2000.
18. Abraham, J.P., Hubert, J.I., George, V., Nielsen, O.F., and Jayakumar, V.S., "Vibrational Spectroscopic Studies on the Natural Product, Columbianadin," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **59**, 193–199, January (2003).
19. Woo, Y.-A., Kim, H.-J., Cho, J.H., and Chung, H., "Discrimination of Herbal Medicines According to Geographical Origin with NIR Reflectance Spectroscopy and Pattern Recognition Techniques," *Journal of Pharmaceutical and Biomedical Analysis*, **21**(2), 407–413 (1999).
20. Woo, Y.-A., Kim, H.-J., Cho, J.H., and Chung, H., "Identification of Herbal Medicines Using Pattern Recognition Techniques with NIR Spectra," *Microchemical Journal*, **63**, 61–70 (1999).
21. Laasonen, M., "Near-Infrared Spectroscopy, a Quality Control Tool for the Different Steps in the Manufacture of Herbal Medicinal Products," *Dissertation Abstracts International*, **64**, 688 (2003).
22. Schulz, H., Schrader, B., Quilitzsch, R., Pfeffer, S., and Krueger, H., "Rapid Classification of Basil Chemotypes by Various Vibrational Spectroscopy Methods," *Journal of Agricultural and Food Chemistry*, **51**, 2475–2481 (2003).
23. Laasonen, M., Harmia-Pulkkinen, T., Simard, C.L., Michiels, E., Raesaenen, M., and Vuorela, H., "Fast Identification of *Echinacea purpurea* Dried Roots Using Near-Infrared Spectroscopy," *Analytical Chemistry*, **74**, 2493–2499 (2002).

24. Kudo, M., Moffat, A.C., and Watt, R.A., "The Rapid Characterization of Natural Products and Herbal Medicines by Near-Infrared Spectroscopy," *The Journal of Pharmacy and Pharmacology*, **50**(Suppl., British Pharmaceutical Conference), 258 (1998).
25. Kudo, M., Watt, R.A., and Moffat, A.C., "Rapid Identification of *Digitalis purpurea* Using Near-Infrared Reflectance Spectroscopy," *The Journal of Pharmacy and Pharmacology*, **52**, 1271–1277, October (2000).
26. Pudelko-Koerner, C., "Quantitative Near-Infrared Reflectance Spectroscopy of Sennosides from *Sennae fructus angustifoliae* in In-Process and Quality Control Including Method Validation," *Pharmazeutische Industrie*, **60**, 1007–1012 (1998).

32 Your Viagras — Genuine, Imitation, or Counterfeit?

*Marjo J. Vredenburg, Dennis Mooibroek, and
Ronald Hoogerbrugge*

CONTENTS

32.1	Introduction	631
32.2	Materials and Methods	632
32.2.1	Samples	632
32.2.2	NIRS Analyses	632
32.2.3	Libraries	632
32.2.4	Chemometrics	632
32.2.5	Chemical Analyses	633
32.3	Results and Discussion	633
32.3.1	Wavelength Correlation	633
32.3.2	Principal Component Analyses	635
32.4	Conclusions	644
	Acknowledgments	645
	References	645

32.1 INTRODUCTION

Since Pfizer introduced (1998) Viagra® tablets for erectile dysfunction, this medicine has been counterfeited and imitated. Counterfeits of Viagra are tablets intentionally manufactured to look like Viagra, but these tablets may contain wrong ingredients or a wrong active substance or no active substance at all. Imitations of Viagra are also manufactured; these tablets do not look like Viagra, but claim or suggest to contain the same active pharmaceutical ingredient of Viagra, sildenafil citrate. Production and distribution of counterfeits and imitations are of great risk for human health [1,6]. Another problem arises with imitation tablets produced in Asia. India and China do not recognize the European patent law, and therefore legally manufactured products such as Kamagra and Edegra are illegal in the Netherlands.

Owing to its discriminating power near-infrared spectroscopy (NIRS) is an attractive analytical technique to investigate the identity of unknown Viagra tablets. NIRS is also very suitable for a large number of individual dosage forms because samples can be measured without any pretreatment and the time of analysis is short. Moreover, the technique is nondestructive.

The power of NIRS to detect counterfeit drugs was investigated by Scafì and Pasquini [2] with a large variety of real samples of counterfeit drugs. These authors reported that their NIRS application was able to recognize all those counterfeit drugs that differed in composition

from the original drug. In addition, Olsen et al. [3] demonstrated the possibilities of NIRS for the screening for counterfeit drugs, using Prozac[®] and its counterfeits as model drugs.

Two chemometric algorithms are performed to process the spectra. Wavelength correlation (WC) is used to distinguish counterfeits and imitations from Viagra tablets and from each other. Principal component analysis (PCA) is applied to compare the spectral information in a large dataset and to detect any clusters of unknown tablets.

The objectives of this study were to develop a quick screening method to distinguish counterfeits and imitations from genuine Viagra, to detect similar batches according to the origin and to the active ingredient and to get insight in possible related sources.

32.2 MATERIALS AND METHODS

32.2.1 SAMPLES

Seven different batches of 50 mg Viagra tablets in closed blisters or jars from European and American makes were kindly donated by Pfizer, the Netherlands. About 120 different batches of unknown tablets (originals, imitations, and counterfeits) were received from the Dutch Health Care Inspectorate. Sample size varied from 1 to 30 tablets. Tablets were delivered in closed or open jars, in blisters, in plastic bags or without any packaging. Samples were confiscated at several places, for example, airport, illegal production places, sex shops, Internet, and so forth. Sildenafil citrate was supplied by Pfizer, the Netherlands. All samples were stored as received in the dark at ambient room temperature and humidity.

32.2.2 NIRS ANALYSES

NIR spectra were recorded on a Spectrum Identicheck Fourier transform near-infrared (FT-NIR) system (PerkinElmer Ltd., Beaconsfield Bucks, UK) with an IdentiCheck Reflectance Accessory (ICRA) with standard Spectrum Identicheck software including WC, version 2.0 to acquire and to process the data. WC was always applied using the default filter setting, which included a resolution filter. Measurements were carried out with an optical resolution of 16 cm^{-1} over the spectral range $12,000$ to $3,000\text{ cm}^{-1}$ and 64 scans were co-added. A PbS detector was used. Spectralon was used as a background reference for solid samples.

All samples were measured as received. NIR spectra were recorded from both sides from at least five tablets (if available) of every batch. Sildenafil citrate (1 g) was measured as delivered in a 4 ml glass vial (Alltech) with closure. The spectra were recorded in the diffuse reflection mode.

32.2.3 LIBRARIES

Two libraries were constructed to screen the unknown tablets. The reference library consisted of ten NIR spectra of five Viagra tablets 50 mg (top and bottom) from each of four batches of European make and three batches of American make. NIR spectra of independent tablets were used to validate the library. The second library consisted of NIR spectra of all measured samples — genuine, imitations, and counterfeits.

32.2.4 CHEMOMETRICS

Screening unknown tablets on authenticity WC was used to compare the NIR spectra ($10,000$ to $4,000\text{ cm}^{-1}$) with the reference library. The correlation coefficient was calculated for every spectrum to determine the similarity with the Viagra tablets. Mean correlation coefficients were calculated from the spectra (top and bottom) of all tablets of one batch. The limit for 'identical' was set at

0.998; higher values indicated that a tablet could not be distinguished from the original Pfizer Viagra tablets. This criterion was used for a single tablet as well as for a whole batch. The similarity with any other NIR spectrum measured before was explored by comparing each spectrum with the second library.

For PCA, all NIR spectra were exported to Matlab (Matlab Release 13, Servicepack 1, August 2003; The MathWorks Inc., Natick, USA). PCA was applied to all 763 NIR spectra in the frequency range 10,000 to 4,000 cm^{-1} without any spectrum pretreatment. The main scoreplot PC1/PC2 and the additional sample information such as trademark, inscription, and appearance were used to detect any clustering in the tablets. A home-made toolbox was developed (Probell Toolbox Version 2.5 June, 2003) to calculate the 95% probability ellipses around the scores of the Viagra spectra and around the most remote clusters in the scoreplot [4]. Owing to the algorithm, it is not possible to draw probability ellipses with less than two points in the figures. Group 1 was always the compilation of all the remaining points (no probability ellipse calculated) and group 10 always consisted of the Viagra reference points.

After removing these most remote clusters from the dataset, a new PCA was calculated with new axis, and it again showed distinguishable clusters. This process was repeated until no further separated clusters could be found. The scoreplots of higher principal components were also investigated to control the consistency of the clusters in other dimensions.

32.2.5 CHEMICAL ANALYSES

After NIRS analysis the presence of sildenafil citrate and other active ingredients was often confirmed. The following three methods were used [5]:

1. A general screening method for pharmaceutical active compounds using HPLC-DAD.
2. Analysis of Sildenafil using DLC and UV/Vis spectrophotometry for identification and UV/Vis spectrophotometry for quantitation.
3. Analysis of Sildenafil for identification and quantitation using HPLC-DAD as described in the registration file of Viagra (Viagra; Dutch registration number: RVG 70340).

32.3 RESULTS AND DISCUSSION

Ten reference spectra of five European Viagra tablets 50 mg (top and bottom) are shown in Figure 32.1. Small absorption bands of the active ingredient, sildenafil citrate, can be seen in the spectral range 6500 to 5500 cm^{-1} (see detail view in Figure 32.1). No differences were found between the reference spectra of European and American makes. The in-between correlation is larger than 0.9980. The NIR spectrum and the molecular structure of Sildenafil citrate is presented in Figure 32.2.

32.3.1 WAVELENGTH CORRELATION

For every batch of unknown tablets the mean correlation coefficient between the spectra of the unknown tablets and the spectra of the reference Viagra tablets was calculated. From 105 different batches the score of only 15 batches exceeded the limit of 0.9980, that is, those samples could not be distinguished from the Viagra tablets. Within these 15 batches three appeared to be original Viagra coming from New Zealand, Australia, and The Netherlands. It was not possible, until now, to determine for the other 12 batches whether these really are extremely well-produced counterfeits (that would mean false positives were found) or that these batches contained originally fabricated Viagra tablets that somehow in the chain from fabrication to retail got lost and entered the illegal circuits.

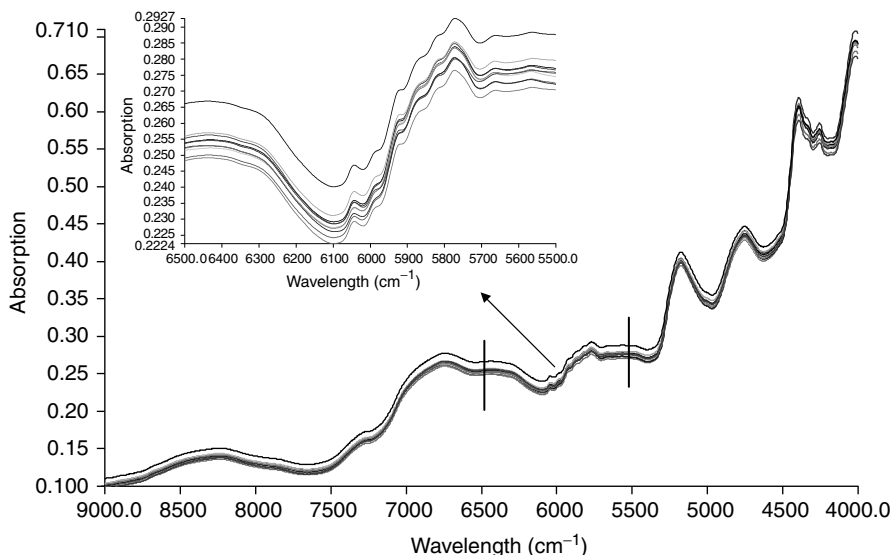


FIGURE 32.1 Ten reference spectra (top and bottom) of five European Viagra tablets 50 mg, and in detail view the absorption bands of the active ingredient Sildenafil citrate.

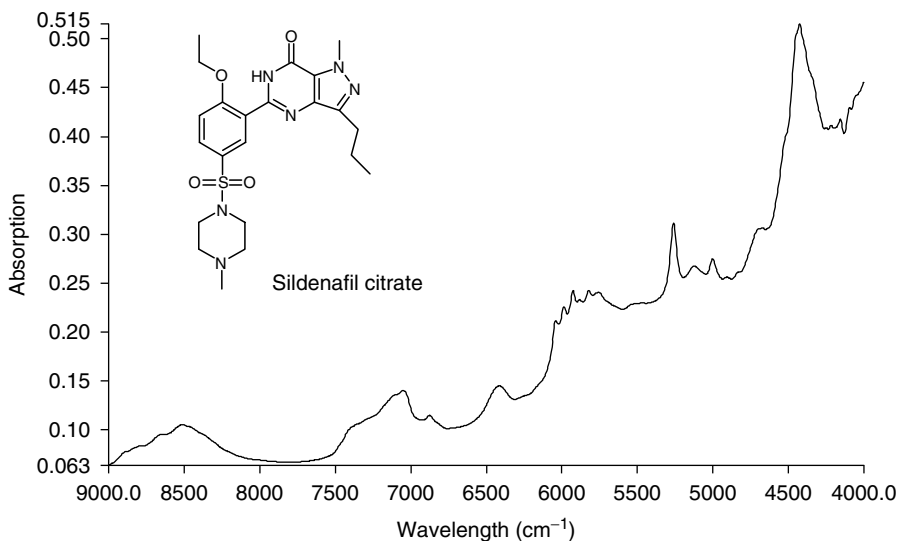


FIGURE 32.2 NIR spectrum of Sildenafil citrate and the molecular structure of Sildenafil.

Another four batches of original Viagra coming from New York, USA, and The Netherlands did not comply with the threshold; however, these batches were delivered in an envelope without the original packaging. A possible explanation is the absorption of water to tablets, but any kind of contamination can occur when tablets are not properly packaged. Apparently, the threshold of 0.998 is applicable only for tablets in original packaging. The library was not validated for unpacked tablets. To avoid rejection of loosely packaged tablets a threshold of 0.995 should be applied. Of course, a lower threshold increases the probability of finding false positives. The first part of the results is shown in Table 32.1. Other results are shown in Table 32.2 to Table 32.7.

TABLE 32.1

The First Part of the Results of the Mean Correlations of Unknown Viagra Tablets with the Reference Library of Viagra Tablets

Group number	Number of tablets	Mean correlation coefficients	Packaging	Trade name, origin
1	10	0.9952	Envelope	Viagra Pfizer, New York
1	10	0.9962	Envelope	Viagra Pfizer, New York
1	10	0.9967	Plastic bag	Viagra Pfizer, USA
1	4	0.9990	Blister	Viagra Pfizer, New Zealand
1	4	0.9971	Blister	Viagra The Netherlands
1	4	0.9991	Blister	Viagra The Netherlands
1	8	0.9994	Blister	
1	4	0.9989	Blister	
1	1	0.9988	Blister	
1	1	0.9988	Blister	
1	4	0.9991	Blister	
1	1	0.9961	Envelope	Internet site
1	1	0.9981	Envelope	Internet site
1	1	0.9985	Envelope	Internet site
1	1	0.9990	Glasvial	Luxembourg
1	2	0.9994	Glasvial	Luxembourg
1	1	0.9995	Blister	UK
1	1	0.9994	Blister	UK
1	1	0.9994	None	Viagra Pfizer, Australia
1	5	0.9992	Jar	

32.3.2 PRINCIPAL COMPONENT ANALYSES

Principal component analysis applied to 763 NIR spectra (step 1) showed a large spread of datapoints in the main scoreplot PC1/PC2 (Figure 32.3). The marked points (group 10) located in the center of the plot are originated from the reference spectra of Viagra tablets, and the dotted line around the group indicates the calculated 95% probability ellipse around the scores. All other points are derived from spectra of other tablets. A large distance of a point to the center group means a great difference compared to the reference spectra. The large diversity of points is caused by the variation in used active ingredients or in the composition of the excipients. Another five clusters (groups 2, 3, 4, 5, and 6) are found after calculating the 95% probability ellipses around the scores of the most remote clusters in the scoreplot. Group 1 represents the remaining points.

The clusters enclosed the spectra of tablets of six different batches from Chinese origin in group 2 containing sildenafil citrate. The separate location in the scoreplot is probably caused by other excipients used for manufacturing. The in-between correlation of these six batches is larger than 0.99. These six batches were part of in-total eight batches, which all were packaged in boxes with different labels. The NIRS analysis revealed that although all labels were different at least some correlation was present in the production/distribution chain of these batches.

Chemical analyses of tablets in three different batches in group 3 demonstrated that the active ingredient was amphetamine. Mean correlation of these samples was 0.7367, 0.8263, and 0.8239, and the NIR spectra are shown in Figure 32.4. These three batches were delivered independently to the laboratory. A trace of caffeine was found in seven different, rather inhomogeneous batches in group 4 confiscated at several places in a single illegal production site. The tablets of one batch from Greece in group 5 probably contained Sildenafil citrate. Four different batches were received from Greece; the tablets had a slightly different color. NIRS analysis indicated that the tablets of these

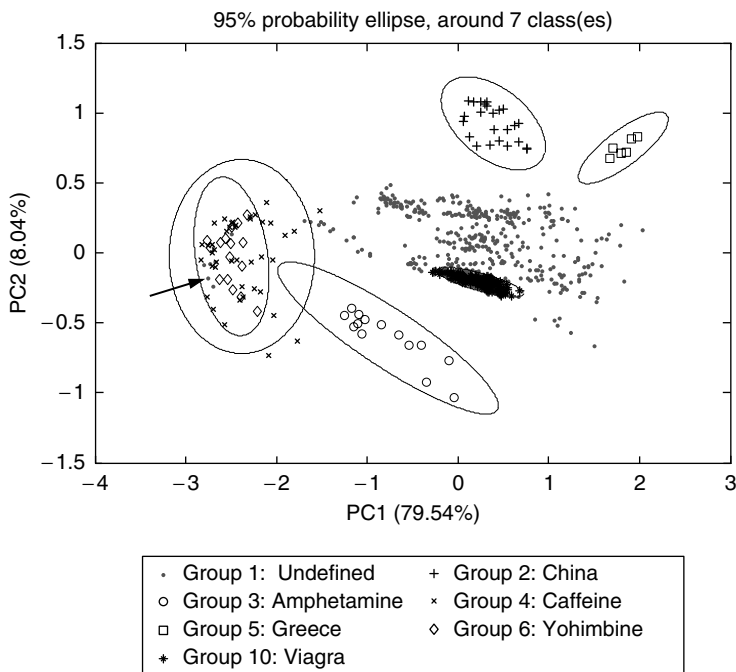


FIGURE 32.3 Scoreplot PC1/PC2 of PCA applied to 763 NIR spectra with 95% probability ellipses.

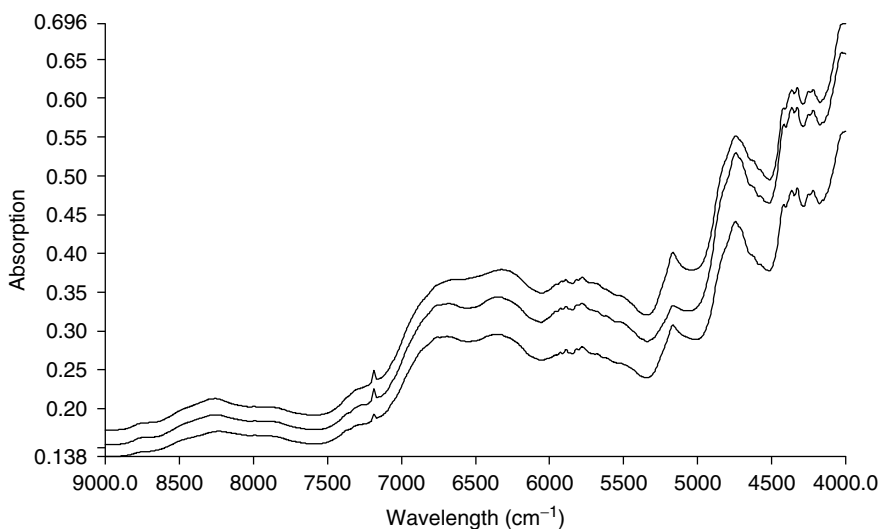


FIGURE 32.4 (See color insert following page 622.) NIR spectra of counterfeit tablets from three different batches containing amphetamine as active ingredient.

batches contained quite different active ingredients. In tablets of two different batches in group 6, NIRS analysis suggested the presence of an unknown active ingredient. Chemical analyses revealed that yohimbine was used as the active ingredient in these tablets. Another batch of three tablets was found to be located in this cluster (see arrow in Figure 32.3); these tablets may, according to the NIR spectra, also contain yohimbine as active ingredient, but confirmation by chemical analysis was not considered necessary.

The corresponding mean correlation coefficients of these batches are shown in Table 32.2. In step 1 a total of 108 spectra of the five clusters mentioned above were removed from the main datafile.

A new scoreplot PC1/PC2 with new axis (Figure 32.5) is shown after performing PCA on the remaining 655 NIR spectra (step 2). The points of the most upper cluster in the center (group 10) belong to the spectra of Viagra tablets. Besides this one, four new clusters (groups 2, 3, 5, and 6) are found after calculating the 95% probability ellipses around the scores of the most remote clusters in the scoreplot. Group 1 represents the remaining points.

The clusters enclosed the spectra of tablets of two batches in group 2 called Androz100, manufactured in India, containing sildenafil citrate and, in contrast to Viagra tablets, an amount of talc. The NIR spectra are shown in Figure 32.6; absorption bands at 7,185 and 10,535 cm^{-1} can be assigned to talc. Furthermore, in one batch in group 3 from Greece, probably containing sildenafil citrate (not confirmed), and one batch in group 5 ordered via Internet and containing sildenafil citrate,

TABLE 32.2

Mean Correlations of All Batches of Unknown Viagra Tablets with the Reference Library of Viagra Tablets Discussed in Step 1

Group number	Number of batches	Tablets per batch	Mean correlation coefficients	Trade name, origin, remarks
2	6	1–3	0.73–0.82	Different trademarks, China
3	3	1–5	0.73–0.84	Containing amphetamine
4	7	3	0.67–0.72	Containing a trace of caffeine
5	1	3	0.82	Greece
6	2	3–5	0.77–0.80	Containing yohimbine

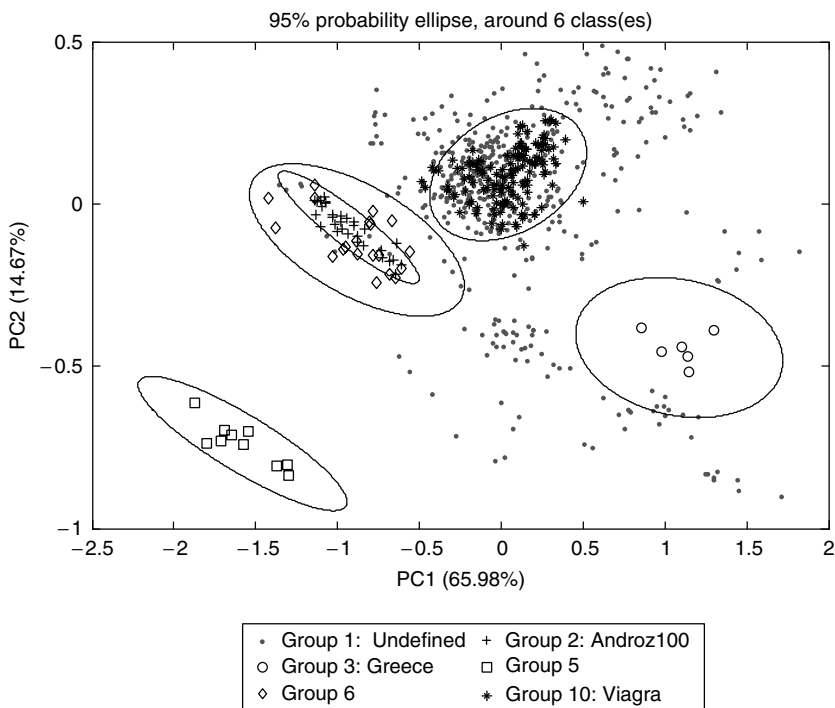


FIGURE 32.5 Scoreplot PC1/PC2 of PCA applied to 655 NIR spectra with 95% probability ellipses.

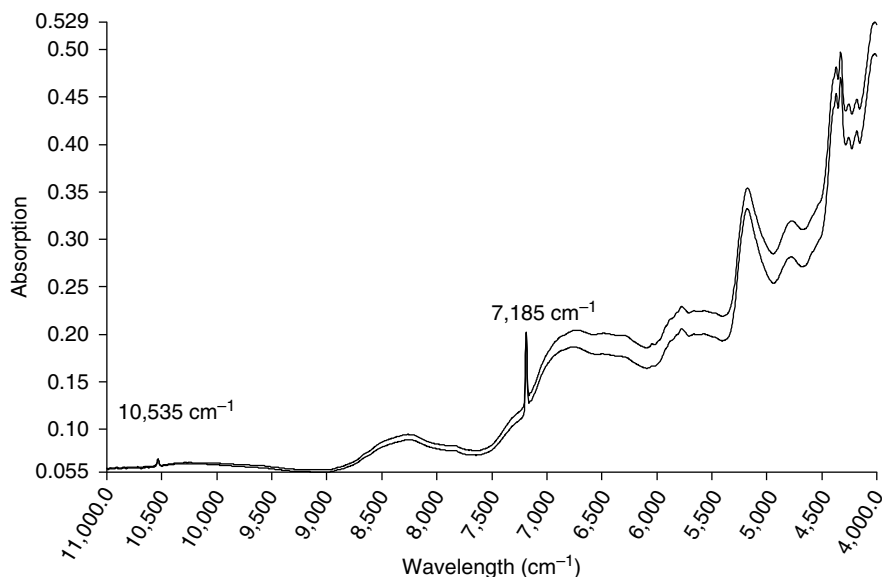


FIGURE 32.6 NIR spectra of Androz100 tablets; absorption bands at 7,185 and 10,535 cm^{-1} indicated an amount of talc.

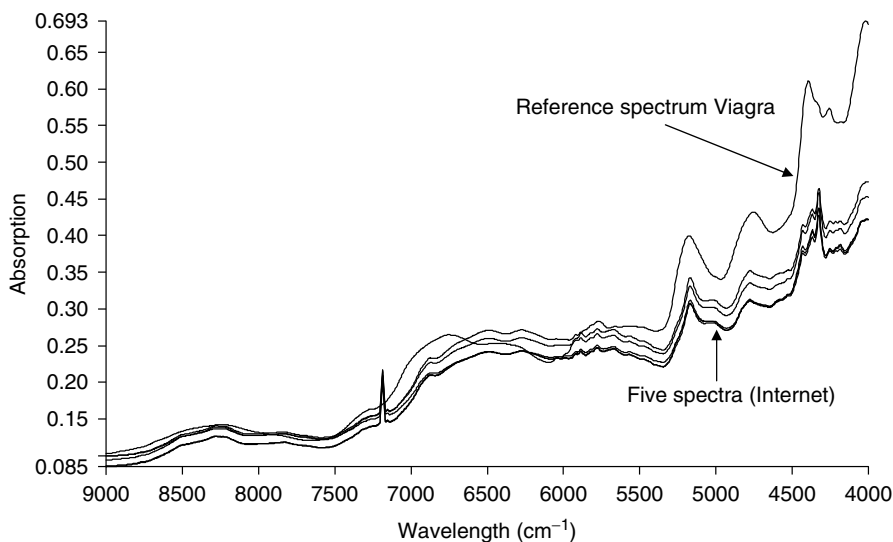


FIGURE 32.7 NIR spectra of five tablets of Viagra 100 mg ordered via Internet and a reference spectrum of Viagra.

tablets were cream-colored inside, which may be the cause for the lowest mean correlation 0.6025 (the spectra are shown in Figure 32.7). The tablets in group 6 are rather inhomogeneous; the 95% probability ellipse — which is much wider than the ellipse around group 2 — overlaps with tablets in group 2. The tablets in both groups are different in color and shape, and in the scoreplots of higher PC's both groups are separated. The corresponding mean correlation coefficients of these batches are shown in Table 32.3. In step 2 a total of 68 spectra were removed from the dataset.

TABLE 32.3

Mean Correlations of All Batches of Unknown Viagra Tablets with the Reference Library of Viagra Tablets Discussed in Step 2

Group number	Number of batches	Tablets per batch	Mean correlation coefficients	Trade name, origin, remarks
2	2	8	0.72–0.73	Androz100, India
3	1	3	0.81	Greece
5	1	5	0.60	Internet site
6	1	10	0.95	

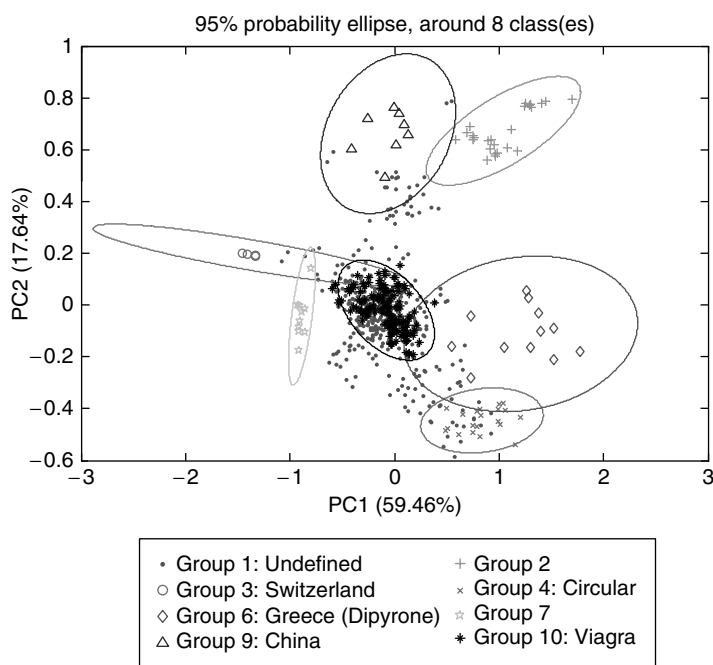


FIGURE 32.8 Scoreplot PC1/PC2 of PCA applied to 587 NIR spectra with 95% probability ellipses.

After performing the PCA on the remaining 587 NIR spectra (step 3) the 95% probability ellipses were calculated around the scores of the most remote clusters in the scoreplot PC1/PC2 (Figure 32.8). Six new clusters (groups 2, 3, 4, 6, 7, and 9) are found.

The clusters enclosed the spectra of tablets of four different batches in group 2 confiscated at different illegal production places containing sildenafil citrate. Although the appearance of the tablets of three batches was different (oblong vs. diamond shape), NIRS analysis revealed that the in-between correlation of these batches was larger than 0.998, indicating an identical chemical composition. Furthermore, one batch in group 3 coming from Switzerland contained no sildenafil citrate or other active ingredients at all. Two different batches of circular blue tablets equally named were found in group 4. These batches were confiscated at different places and the tablets were packaged in gifts and contained sildenafil citrate. Although the gifts were rather different NIRS analysis showed a high correlation >0.99 between the tablets in these batches. Two batches from Greece were found in group 6, chemical analyses demonstrated that both contained dipyron as active ingredient, one batch in group 7 containing sildenafil citrate and one of two batches of Chinese make in group 9

consisted of brownish pink colored tablets. Also the spectra of one single blue-spotted oblong tablet containing sildenafil citrate was found outside the probability ellipse of Viagra tablets.

The corresponding mean correlation coefficients of these batches are shown in Table 32.4. In step 3 a total of 81 spectra were removed from the dataset.

After performing PCA on the remaining 506 NIR spectra (step 4) the new scoreplot PC1/PC3 is shown in Figure 32.9. In the scoreplot PC1/PC2 points of group 3 are overlapping with group 10 (Viagra tablets); however, in the scoreplot PC1/PC3, the groups are fully separated and five new clusters (groups 2, 3, 5, 6, and 7) are found after calculating the 95% probability ellipses. Group 1 represents the remaining points.

TABLE 32.4

Mean Correlations of All Batches of Unknown Viagra Tablets with the Reference Library of Viagra Tablets Discussed in Step 3

Group number	Number of batches	Tablets per batch	Mean correlation coefficients	Trade name, origin, remarks
2	4	1–7	0.89–0.94	Unexpected identical composition
3	1	2	0.95	Switzerland
4	2	5	0.92–0.93	Circular blue tablets in gifts
6	2	3	0.81–0.89	Containing dipyrone, Greece
7	1	5	0.99	
9	2	2	0.93–0.94	Different trademarks, China
	1	1	0.91	Blue-spotted oblong tablet

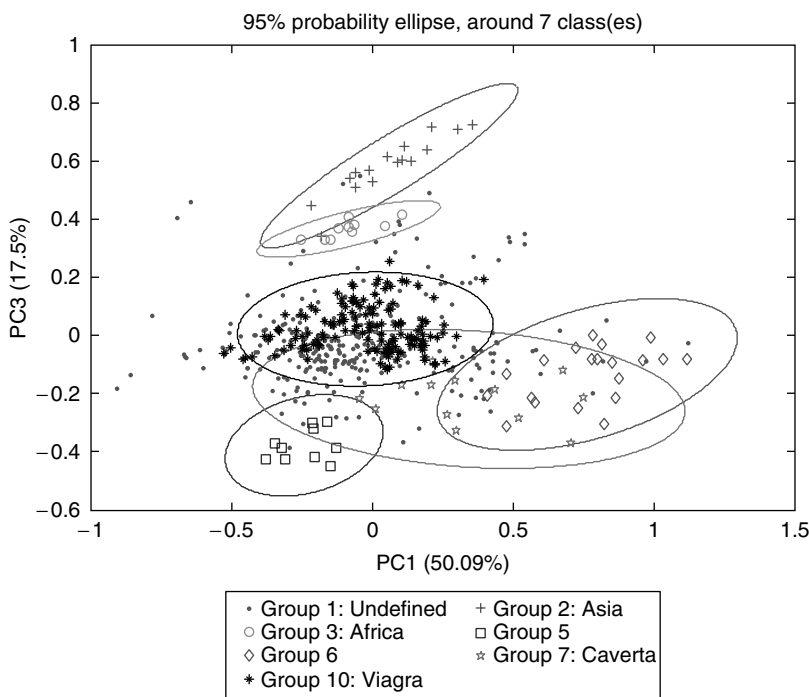


FIGURE 32.9 Scoreplot PC1/PC3 of PCA applied to 506 NIR spectra with 95% probability ellipses.

TABLE 32.5

Mean Correlations of All Batches of Unknown Viagra Tablets with the Reference Library of Viagra Tablets Discussed in Step 4

Group number	Number of batches	Tablets per batch	Mean correlation coefficients	Trade name, origin, remarks
2	1	8	0.89	Asia
3	1	5	0.95	Africa
5	1	5	0.93	
6	2	5	0.91	Unexpected identical composition
7	5	1–2	0.96–0.98	Caverta, India
	1	1	0.96	
	1	1	0.92	

The clusters enclosed the spectra of tablets of one batch in group 2 containing sildenafil citrate and coming from Asia confiscated at the airport, one batch in group 3 containing sildenafil citrate and coming from Africa also confiscated at the airport and one batch in group 5 containing sildenafil citrate and two batches in group 6 containing sildenafil citrate confiscated at different places. NIRS analysis indicated a high similarity; the in-between correlation was larger than 0.998 and the tablets of these batches may come from the same production site. Group 7 consisted of five different batches of dark red tablets with a triangle shape both 50 and 100 mg called Caverta, all containing sildenafil citrate. Also, the spectra of two different single tablets not belonging to any group were found outside the probability ellipse of Viagra tablets.

The corresponding mean correlation coefficients of these batches are shown in Table 32.5. In step 4 a total of 73 spectra were removed from the dataset.

After performing the PCA on the remaining 433 NIR spectra (step 5) the 95% probability ellipses were calculated around the scores of the most remote clusters in the new scoreplot PC1/PC2 (Figure 32.10). Six new clusters (groups 2, 3, 4, 5, 6, and 7) are found. Some clusters with only a few points had a large probability ellipse around the points; for viewing purposes these large ellipses are not visible in the presented plot.

The clusters enclosed the spectra of tablets of one batch in group 2 containing sildenafil citrate and coming from Asia confiscated at the airport, one batch in group 3 of Chinese make containing sildenafil citrate, one batch of circular blue tablets named Edegra in group 4 containing sildenafil citrate, and one batch in group 5 containing sildenafil citrate. Two batches in group 6 confiscated at different places contained sildenafil citrate, and the correlation coefficient >0.998 indicated an identical chemical composition. Furthermore, one batch of diamond-shaped red tablets containing sildenafil citrate was found in group 7 confiscated at an illegal production place. Again, the spectra of three different single tablets not belonging to any group were found outside the probability ellipse of Viagra tablets.

The corresponding mean correlation coefficients of these batches are shown in Table 32.6. In step 5 a total of 62 spectra were removed from the dataset.

A new scoreplot PC1/PC2 (Figure 32.11) is shown after performing PCA on the remaining 371 NIR spectra (step 6). Seven new clusters (groups 2, 3, 4, 5, 6, 8, and 9) are found after calculating the 95% probability ellipses around the scores of the most remote clusters in the scoreplot. Some clusters with only a few points had a large probability ellipse around the points; for viewing purposes these large ellipses are not visible in the presented plot. Another new cluster (group 11) is found in the scoreplot PC1/PC3. Group 1 represents the remaining points.

The clusters enclosed the spectra of tablets of one batch of circular blue tablets named Edegra in group 2 containing sildenafil citrate and five different batches with high similarity in group 3 containing sildenafil citrate. The in-between correlation of these batches is greater than 0.998. For

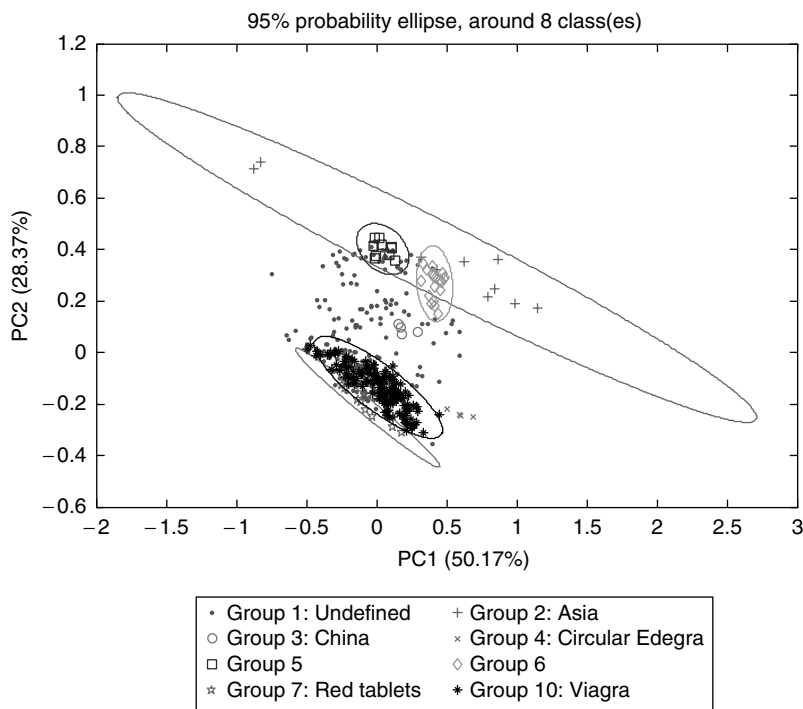


FIGURE 32.10 Scoreplot PC1/PC2 of PCA applied to 433 NIR spectra with 95% probability ellipses.

TABLE 32.6

Mean Correlations of All Batches of Unknown Viagra Tablets with the Reference Library of Viagra Tablets Discussed in Step 5

Group number	Number of batches	Tablets per batch	Mean correlation coefficients	Trade name, origin, remarks
2	1	5	0.91	Internet site
3	1	2	0.94	China
4	1	2	0.93	Edegra, India
5	1	5	0.93	
6	2	5	0.94	Unexpected identical composition
7	1	4	0.99	Diamond-shaped red tablets
	1	1	0.95	
	1	1	0.96	
	1	1	0.89	

ten inhomogeneous batches in group 4 confiscated at the airport, NIRS revealed that each batch contained two types of tablets with different active ingredients. Chemical analyses indicated the presence of clomiphene citrate or sildenafil citrate. Criminal investigation confirmed that these tablets originated from the same source. Furthermore the clusters also enclosed one batch in group 5 manufacture Kamagra containing sildenafil citrate, one batch in group 6 containing sildenafil citrate, one batch in group 8 ordered via Internet containing sildenafil citrate, one batch in group 9 containing sildenafil citrate, and one batch in group 11 containing sildenafil citrate. The spectra of one single tablet not belonging to any group were found outside the probability ellipse of Viagra tablets.

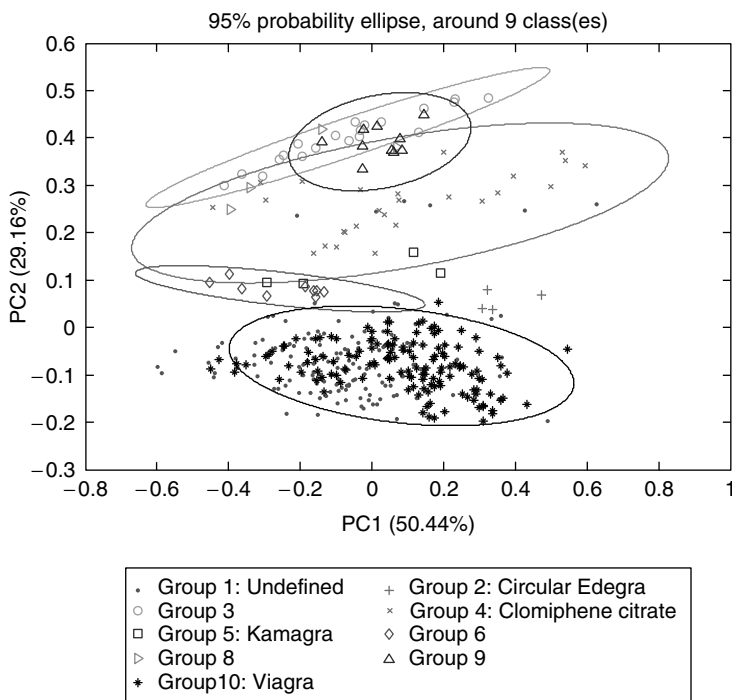


FIGURE 32.11 Scoreplot PC1/PC2 of PCA applied to 371 NIR spectra with 95% probability ellipses.

TABLE 32.7

Mean Correlations of All Batches of Unknown Viagra Tablets with the Reference Library of Viagra Tablets Discussed in Step 6

Group number	Number of batches	Tablets per batch	Mean correlation coefficients	Trade name, origin, remarks
2	1	2	0.93	Edegra, India
3	5	2	0.94	
4	10	2–3	0.95–0.97	Inhomogeneous ^a
5	1	4	0.98	Kamagra, Internet site
6	1	9	0.98	
8	1	2	0.94	Internet site
9	1	5	0.94	
11	1	2	0.97	
	1	1	0.97	

^a Inhomogeneous batches containing clomiphene citrate or sildenafil citrate.

The corresponding mean correlation coefficients of these batches are shown in Table 32.7. In step 6, a total of 79 spectra were removed from the dataset.

After performing the PCA on the remaining 292 NIR spectra, no more separated clusters could be detected in the scoreplots PC1/PC2 (Figure 32.12) and further. The points of the remaining spectra in group 1 overlap with the marked points from the reference spectra of Viagra tablets in group 10.

Group 1 enclosed seven batches of original Viagra tablets coming from New York, USA, New Zealand, The Netherlands, and Australia. Five batches of unknown tablets and eight single tablets

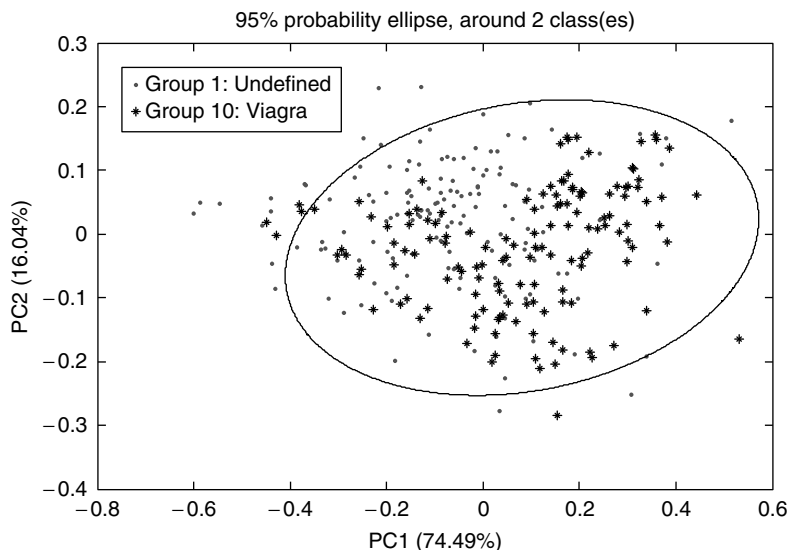


FIGURE 32.12 Scoreplot PC1/PC2 of PCA applied to 292 NIR spectra with 95% probability ellipses.

could not be distinguished from Viagra tablets. The corresponding correlation coefficient calculated with WC can be seen in Table 32.1.

The results of removing clusters from the main dataset PCA completely match with the results of WC. As regards the Table 32.2 to Table 32.7 comparing spectra of unknown tablets with the reference library, an increase in correlation coefficients can be seen, meaning that the distance to the cluster of reference spectra is decreasing.

32.4 CONCLUSIONS

Near-infrared spectroscopy combined with WC and PCA has proven to be a good technique for screening unknown Viagra tablets on authenticity. On the basis of the reproducibility of spectra from a large number of Viagra tablets from different batches, an extremely sharp WC threshold of 0.998 appeared to be applicable. From 105 batches of “unknown” tablets, usually of illegal origin, 86 batches (82%) did not meet the threshold and were identified as imitation or counterfeit. Fifteen batches did exceed the threshold. Within these 15 batches, 3 batches appeared to be genuine Viagra. These batches delivered in original packaging (closed blisters or jars) are correctly identified. For the remaining twelve batches it is until now not clear whether they are extremely well-fabricated counterfeits or an original fabricate that somehow entered illegal circuits. Four batches of genuine Viagra tablets without proper packaging did not comply with the criterion, perhaps due to the influences of water absorption.

Within the set of all batches analyzed a remarkable amount of substructure is presented using PCA with a stepwise removal of the most remote clusters. Some of these clusters correlate with the application of another active ingredient. For example, several batches containing amphetamine were discovered. The resemblance between these batches was not expected on forehand. The presence of amphetamine was predicted by using the NIR spectra, and it was chemically confirmed. Similar phenomena were observed for yohimbine as active ingredient. In general, the quality of imitations, yielding compact clusters, was much better than the quality of counterfeited tablets confiscated at illegal production places. The latter have such large intra batch variability that the assignment of related sources is sometimes difficult.

Finally, the NIR spectra provide a lot of information about the composition, the homogeneity, the quality of a batch, and all kind of unexpected variations in the unknown tablets. This additional information is most valuable to guide further chemical and forensic analyses.

ACKNOWLEDGMENTS

The authors thank Dr D. de Kaste of Centre for Quality of Chemical-Pharmaceutical Products of the RIVM under whose authority this research was performed and the colleagues of Centre for Quality of Chemical-Pharmaceutical Products for carrying out all chemical analyses.

REFERENCES

1. British Pharmaceutical Conference 2003; session counterfeiting, *The Pharmaceutical Journal* (2003), 271: 453–454.
2. S.H. Scafì and C. Pasquini, Identification of counterfeit drugs using near-infrared spectroscopy, *Analyst* (2001), 126: 2218–2224.
3. B.A. Olsen, M.W. Borer, F.M. Perry, and R.A. Forbes, Screening for counterfeit drugs using near-infrared spectroscopy, *Pharmaceutical Technology North America* (2002), 26: 62–71, 95.
4. J. Edward Jackson, *A User's Guide To Principal Components*, John Wiley & Sons, New York, 1991. Chapter 15 Flatland: Special Procedures for Two Dimensions 342–349.
5. M.J. Vredenbregt, L. Blok-Tip, R. Hoogerbrugge, D.M. Barends, and D. de Kaste, Screening suspected counterfeit Viagra and imitations of Viagra with Near Infrared Spectroscopy, *Journal of Pharmaceutical and Biomedical Analysis* (2006), 40: 840–849.
6. L. Blok-Tip, H. Vogelpoel, M.J. Vredenbregt, D.M. Barends, and D. de Kaste, Counterfeits and imitations of Viagra and Cialis tablets: trends and risks to public health, RIVM report 267041001/2005, Bilthoven, The Netherlands.

33 Biomedical Applications of Near-Infrared Spectroscopy

Emil W. Ciurczak

CONTENTS

33.1 Blood Glucose	647
33.2 Blood Oxygen.....	648
33.3 Tissue.....	650
33.3.1 Chemistry (<i>In Vitro</i>)	652
33.3.2 (Bio) Chemistry (<i>In Vivo</i>)	652
33.3.3 Cancer	653
33.4 Physics, Physical Parameters, Math, and Imaging	653
References	653

While the various applications in this chapter are listed by topic — blood chemistry, blood oxygen, and so forth — there may be overlaps or applications where a single application is difficult to assign.

33.1 BLOOD GLUCOSE

One of the most publicized and pursued uses of near-infrared (NIR) in the life sciences is for *in situ* glucose measurements. A number of patents for noninvasive blood-glucose devices have been issued. For instance, the one developed by Ham and Cohen [1], patent number 5,553,616, where the light is passed through the finger. The poster, by Gabriely et al. [2] acquired spectral data from the thumb and used it to measure clinically relevant plasma glucose scanning from 400 to 1700 nm with a fiber-optic probe.

Modeling of the glucose-blood system has been an ongoing project for Gary Small (Ohio University) and Mark Arnold (University of Iowa). In 1993 [3], they published a paper modeling glucose in a protein-containing matrix. They found a partial least-squares (PLS) regression gave the best correlation with a standard error of 0.24 mM.

In 1994, they reported a temperature-insensitive method [4] where temperatures between 32 and 41°C were investigated. The variations caused large variations in the spectra from water-band shifts. Fourier-filtering effectively eliminated these differences (standard error of estimate [SEE] = 0.14 mM).

Small and Arnold et al. (1996) [5] used physiological levels of glucose (1 to 20 mM) in the presence of protein and triglycerides. Multivariate manipulation algorithms compensated for the variations in the blood as the glucose level remains unchanged. They also published work on using quadratic-PLS and digital-filtering techniques to account for nonglucose-related changes in the spectra [6].

Two papers by Arnold et al. (1998) [7,8], were devoted to the calculations used in noninvasive blood-glucose measurements. They used neural networks (NN) and PLS to understand how light is attenuated as it passes through tissue. They recommended that fat, water, and tissue be compensated for in any model considered.

David Haaland et al. [9] published a paper in 1992 using whole blood for the model. Scanning from 1500 to 2400 nm, a PLS equation was developed on glucose-spiked whole blood. The range of 0.17 to 41.3 mM gave an equation with a SEE = 1.8 mM.

A device by Schrader [10] uses a laser to illuminate the humor of the eye to measure blood glucose, based on a patent by Schrader et al. [11]. Glucose levels in the anterior chamber of the eye were measured with a latency of approximately 20 min. The equations developed had an error of ± 30 mg/dl.

NIR through skin and muscle shows the blood-glucose level is similar to the glucose level in tissues [12]. Contradicting this work was a paper by Sternberg et al. [13]. They claimed that tissue contained only 75% of the level found in the blood. Most readings in NIR are inclusive of blood and tissue.

Arnold et al. [14] built phantoms of water, fat, and muscle tissue reading the skin of a patient. *In vivo* overtone spectra collected across human webbing tissue (6.7 mm thick) are simulated with water-layer thickness from 5.0 to 6.4 mm (when combined with fat-layer thickness from 1.4 to 4.2 mm).

This “phantom” work was continued (Arnold et al.) in a later publication [15], designed to “warn” the inexperienced user of the pitfalls of chemometrics. They carefully omitted glucose from their samples that were randomly assigned glucose values and a PLS regression was performed. An equation was developed that gave reasonable standard errors, regression coefficients, and so forth. However, this equation could not predict glucose when actual samples were tested.

Maier et al. [16] observed correlated blood glucose with reduced scattering coefficient of tissue. The scattering coefficient was measured with a high enough precision to detect changes in glucose.

Researchers at the University of Krakow published several math treatments of the spectra produced by blood through skin and muscle [17–19]. They used NN, however, these algorithms could not be run on desktop computers a decade ago.

One technique [20], used a fiber-optic lightpipe through the skin of a finger. The device used a portion of fiber, stripped of cladding, as an attenuated total reflectance (ATR) cell. Much of the radiation is lost in the skin, so white light was used. The monochromator was post-sample, giving better sensitivity.

Heise et al. [21] published a procedure for measuring blood glucose through the lip. At 1100 to 1800 nm, using PLS, the mean-square prediction error ($\text{PRESS}^{1/2}$) was between 45 and 55 mg/dl.

33.2 BLOOD OXYGEN

One early report of NIR for diagnostic applications came from Jobsis in 1977 [22]. He monitored the degree of oxygenation of certain metabolites. Later, Ozaki et al. [23] examined venal blood to determine the level of deoxyhemoglobin. Using a miniature integrated-sphere, the back of the hand was illuminated. The 760-nm band in the spectrum correlated well with deoxyhemoglobin.

Michael Sowa [24] used NIR imaging to monitor regional and temporal variations in tissue oxygenation. He studied effects of periods of restricted blood outflow (venous outflow restriction) and interrupted blood inflow (ischemia). Multivariate analyses of image (Fuzzy C-means) and spectral data time courses were used to identify correlated spectral and regional domains. The region from 400 to 1100 nm was plotted as a “topographical” representation of the phenomenon. Peaks and valleys were apparent where blood became oxygenated and deoxygenated.

Mancini et al. [25] estimated skeletal muscle-oxygenation by using the differential absorption properties of hemoglobin. Oxygenated and deoxygenated hemoglobin have identical absorptivities at 800 nm, while deoxygenated hemoglobin predominates at 760 nm. Venous oxygen saturation and absorption between 760 and 800 nm were correlated.

The influence of fat layers on measuring the oxygenation of blood was examined by Lin et al. in 1998 [26]. The phantom experiments showed fat makes a difference in patient-to-patient measurements. These may be compensated in any individual patient. Yamamoto [27] addressed the issue of fat interference with an oximeter that corrected for the influence of subcutaneous fat.

The water effects on hemoglobin concentration in a tissue-like phantom were studied by Franceschini et al. in 1996 [28]. DiMarzio's students [29] at Northeastern University built a device to measure blood-oxygen levels in the brain, particularly for newborns. A NIR beam enters the patient's head at one point and a second probe is used to collect light at a second point.

Jiang et al. [30] presented a device to measure cerebral tissue-oxygenation using fiber-optics and shorter wavelengths. A device to perform diffuse reflectance measurements on the skin was developed by Marbach and Heise [31]. It has an on-axis ellipsoidal collecting mirror with efficient illumination for small sampling areas of bulky body specimens.

Keiko Miyasaka presented work [32] at a meeting held in Toronto. He introduced what he calls a "Niroscope" for NIR spectroscopy. His work was performed during pediatric anesthesia and intensive care. He found Beer's law was not followed rigorously when the signal was passed through the cranium.

What Miyasaka was measuring was the intercranial chromopher levels of oxygenated hemoglobin (HbO_2), deoxygenated or reduced hemoglobin (Hb), and cytochrome redox status. Two methods were used: photon counting and a micro-type pulse laser. Van Huffel et al. [33,34] from Belgium monitored the brain to correlating behavioral states of preterm infants to understand the development of brain-hemodynamics autoregulation. The concentrations of HbO_2 , Hb, and Cytochrome aa₃ (Cytaa₃) were used to monitor the oxygenation level in infant brain blood.

Cooper et al. in 1998 [35] performed another study on the adult brain. NIR was used to determine the effects of changes in the rate of oxygen delivery on adult rat-brain chemistry. Absolute levels of oxyhemoglobin, deoxyhemoglobin, and the redox state of the Cu_A center in mitochondrial cytochrome oxidase. A study on human infants was performed by Wyatt et al. [36]. They used NIR to quantify the cerebral blood volume in human infant.

Kupriyanov et al. [37] determined intracellular pO₂ in cardiac muscle by the balance between its diffusion from vascular (VS) to intercellular space (IS) and its uptake by mitochondria. Ischemia in the forearm was studied by Mansfield et al. in 1997 [38]. In this study, the workers used Fuzzy C-means clustering and principal component analysis (PCA) of time series from the NIR imaging of (volunteers') forearms. They attempted predictions of blood depletion and increase without a priori values for calibration.

Wolf et al. in 1996 [39] used NIR and laser Doppler flowmetry (LDF) to study the effect of systemic nitric oxide synthase (NOS) inhibition on brain oxygenation. The study, performed on rats, demonstrated no effects on brain oxygenation during cortical spreading depression (CSD).

Doppler ultrasound was combined with NIR in a study by Liem et al [40]. They followed the cerebral oxygenation and hemodynamics in preterm infants treated with repeated doses of indomethacin. In addition to the normal concentrations of oxyhemoglobin, deoxyhemoglobin, and oxidized cytochrome aa₃ measured by NIR, transcutaneous pO₂ and pCO₂, arterial O₂ saturation, and blood pressure were measured as well.

A piece of equipment, developed to measure oxygen content, was produced by the Centre for Biomedical Technology in Australia [41]. It consisted of five 1W lasers at wavelengths of 780, 800, 830, 850, and 980 nm and used a photodiode receiver. It used the hemoglobin/deoxyhemoglobin absorbance differences and measured the SO₂ content of the blood as well.

The physical placement of detectors on the scalp for brain blood oxygenation was studied by Germon et al. [42] in a 1998 study. Detectors placed 2.7 and 5.5 cm from an NIR emitter were compared for the determination of Hhb, O₂Hb, oxidized cytochrome C oxidase, and total hemoglobin. The signal change/photon pathlength detected at 5.5 cm was significantly greater for Hhb than for 2.7 cm. The increase in all chromophores detected at 5.5 cm during scalp hyperemia was significantly less than at 2.7 cm.

With similar instrumentation, Henson et al. [43] determined the accuracy of their cerebral oximeter under conditions of isocapnic hypoxia. Using healthy volunteers, dynamic end-tidal (ET) forcing was used to produce step changes in $P_{ET}O_2$, resulting in arterial saturation ranging from approximately 70 to 100% under conditions of controlled normocapnia (resting $P_{ET}O_2$) or hypercapnia (resting plus 7 to 10 mmHg). The O_2 concentrations for each patient under each condition were determined: (a) Hoshi et al. [44] investigated the neuronal activity, oxidative metabolism, and blood supply during mental tasks; (b) Okada [45] presented work on impaired interhemispheric integration in brain oxygenation and hemodynamics in schizophrenia; (c) Hoshi [46] looked into the features of hemodynamic and metabolic changes in the human brain during all-night sleep; and (d) Hirth et al. [47] studied the clinical application of NIR in Migraine patients. They assessed the transient changes of brain tissue oxygenation during the aura and headache phases of a migraine attack.

Surgeons are concerned with brain blood flow to patients undergoing cardiopulmonary bypass surgery. An intensive study by Chow et al. [48] was conducted where blood flows were restricted to patients from age 2 weeks to over 20 years. NIR was used to correlate blood-flow rate with NIR spectra of the brain. Flows of 0.6, 1.2, and 2.4 L/m²/min were used. Their results showed that flow was related to mean arterial pressure, but did not correspond to pulsatility. Pulse rate is often used as a diagnostic to assure sufficient blood flow to the brain during surgery.

Totaro et al. [49] published a paper on the factors affecting measurement of cerebrovascular reactivity when measured by NIRS. They covered the relative transparency of the skin, skull, and brain in the 700 to 1100 nm region and the oxygen-dependent tissue absorption changes of hemoglobin. The test was based on a three-min baseline, a three-min hypercapnia (5% CO_2 in air), and a two-min recovery period.

Changes in NIR spectra and transcranial Doppler sonography parameters were significantly correlated with variations of end-tidal CO_2 ($P < .005$). Their overall conclusion was that NIR was a viable technique for evaluation of cerebrovascular reactivity for patients with cerebrovascular disease.

Exciting work was reported by Hitachi at a meeting in Japan [50]. The research used NIR to detect blood flow changes in the brain to determine sites of epileptic activity. The location of blood flow increases responded well with conventional methods such as intercranial electroencephalogram (EEG) or Single Photon Emission Computed Topography (SPECT). The technique was able to determine the side of the brain where the episode was taking place in all the patients on which it was tried.

33.3 TISSUE

Dreassi et al. [51] published a series of papers on atopy of skin. The first, discusses how NIR penetrates complex structured matrices to at least 0.20 mm. He found NIR to give valuable insight into the stratum corneum. Using PCA, they decomposed the global structural information into components such as water and lipid structures. The group also studied interactions between skin and propylene glycol 400 (PEG 400), isopropyl myristate (IPM), and hydrogel [52]. They examined spectral differences and differences in response in terms of water and lipid content between normal and atopic skin after reaction with these reagents.

Using PCA, they were able to distinguish atopic from normal skin simply from contact with these reagents. Similar results were reported in a later work from this group [53] (third in a series). Using a series of perfluorinated polyethers (fomblins) of differing molecular weight and viscosity, the NIR spectra of normal and atopic skin were assessed.

The interaction between the chemicals and the skin consist of two stages; it is physically modified and the water is moved and redistributed. One important assessment made with NIR is the viability of tissue after trauma [54]. Prolonged and severe tissue hypoxia results in tissue necrosis in pedicled flaps. The group used NIR to identify tissue regions with a low oxygen supply. It was seen that oxygen delivery to the flap tissue dropped immediately. As expected, severe trauma that causes

severing of the skin from the main blood flow causes necrosis of the tissue. NIR may be used as a tool in assessing the success of reattachment of the traumatized skin.

Similar work is being performed by Folke Lind of Sweden [55]. He is introducing monitoring equipment to monitor tissue oxygenation of patients in need of hyperbaric oxygen (HBO) treatment. Since hypoxia is reversible, NIR is believed to be the most rapid tool in a trauma situation.

Fingernails were studied by Sowa et al. [56] both *in vivo* and *ex vivo*. Mid-IR (MIR) and NIR spectra were taken of viable and clipped human nails. Depth profiling by MIR was physically performed while performed noninvasively by photoacoustic spectroscopy (PAS), NIR-ATR, NIR diffuse reflectance, and PAS were compared. Assignments were made, such as N—H stretch-amide II bend combination being centered at 4868 cm^{-1} in this basic study. They concluded that the lower energy NIR-ATR, for purposes of their study, gave the best results.

Measurements such as body fat in infants are easily made by NIR [57]. Newborns can be evaluated for body fat due to breast feeding vs. non-breast feeding. This is more accurate than body weight that includes other tissues and bone.

Another application to pre- and newborns was published by Liu et al. in 1997 [58]. NIR was used as a measure of fetal lung maturity from the spectra of amniotic fluid. The lecithin/spingomyrlin (L/S) ratio was determined by thin layer chromatography (TLC) and used to calibrate a NIR equation using whole amniotic fluid extracted from pregnant women. About $350\text{ }\mu\text{l}$ of fluid was required. This was scanned from 400 to 2500 nm. The correlation between further samples of fluid and TLC results was about 0.91. Of course, a PLS equation was needed due to the complex nature of the samples.

Tissue temperature was measured by Barlow et al. in 1995 [59]. Absorbance changes in the transmission water spectrum between 700 and 1600 nm and between 800 and 2200 nm (reflectance) were found to correlate with the temperature of the tissue. The SEE = 0.02 to 0.12°C and SEP = 0.04 to 0.12°C .

Tissue is a highly dispersing medium and various attempts have been made to mitigate this scattering. Tsai et al. [60] presented a paper concerned with the absorption properties of soft-tissue constituents. They used the region from 900 to 1340 nm. Shorter NIR wavelengths have lower scattering coefficients, lower absorptivities, and, consequently, deeper penetration in tissue. At the same symposium, Schmitt et al. [61] presented a paper discussing the processing of spectra from turbid biological tissue.

Because tissue both scatters and attenuates any light passing through it, many papers have been published about the measurement approach. Rava et al. [62] used NIR Raman to generate spectra, using a Nd:YAG laser to penetrate the tissue with sufficient power. They used a charge-coupled device (CCD) to collect sufficient light for a high signal-to-noise spectrum.

Anderson-Engels et al. presented a series of papers on time-resolved transillumination of tissue, specifically with tumor detection in mind [63–67]. In these paper, the group details the physics involved in using a pico-second diode-laser, a mode-locked argon-ion/dye-laser, or a mode-locked Ti:sapphire laser to conduct time-resolved spectroscopy on tissue. In one case, a human (female) breast is compressed to approximately 35 mm for the test. Light, in 100 femtosecond (fs) pulses (at 792 nm) (giving a 50 ps apparatus function), are dispersed to a signal that is more than 1 ns long. The dispersion curve obtained contains information about the optical properties of the tissue. In the case of scattering-dominated attenuation (scattering coefficient \gg absorption coefficient) detection of early transmittal light will be practically insensitive to variations in the absorption coefficient. The scattering properties determine the amount of detected early light.

Robert Lodder [68] has been producing excellent results in the assessment of arterial walls for a decade. In this, the location and quantities of HDL, LDL, and apolipoproteins in living tissue were determined. A compound parabolic concentrator (CPC) was used to compress the beam from a transmitting optical fiber onto a small spot on the artery surface. Light in the 1100 to 2500 nm range was transmitted through the concentrator onto the exterior arterial wall. The scattered light was detected at the proximal end of the CPC by lead sulfide detectors, located off axis of the incident beam. False color maps are then produced wherein the types and amounts of each type of plaque are determined.

Recently, Lodder's group presented the latest software breakthrough [69]. The procedure predicted plaque at risk of breaking free and causing potential strokes. The software, named CALDATAS, was the latest multivariate program developed by Lodder, beginning with two called BEAST and BEST. It is recommended that his Web page be visited for a full accounting of the work [70,71].

33.3.1 CHEMISTRY (*IN VITRO*)

Cell culture media was analyzed by McShane and Cote' in 1998 [72]. Samples of a three-day fibroblast culture were analyzed by standard clinical techniques as well as by NIR. Glucose, lactate, and ammonia were determined after building a model from several lots of cell culture media. The purpose was to follow the nutrient levels to determine noninvasively when fermentation was complete. A study by Jeff Hall [73] was performed, using NIR to analyze the major components of human breast milk.

Shaw et al. [74] performed analyses of urine samples. They quantified protein, creatinine, and urea. The SEP (using PLS) for the urea, creatinine, and protein were 16.6, 0.79, and 0.23 mmol/l, respectively. Further urine analyses were performed by Jackson et al. in 1977 [75]. Urine glucose, protein, urea, and creatinine concentrations were analyzed using simple algorithms. Urea was calibrated by correlating the absorbance at 2152 nm. Comparison with standard methods gave linearity with a slope of approximately 1.00.

As an example of *ex vivo* determinations, Shaw et al. [76] were able to correlate the chemistry of Synovial fluid, drawn from patients' knees, with arthritis diagnosis. A model equation was built using PLS. The prediction of further arthritis sufferers was remarkably good when the equation was tested.

Schultz et al. [77] used NIR and MIR to examine the structure of Ribonuclease A (RNase A). In MIR, the thermal unfolding of the protein was used as a model for the structural changes occurring in water. In NIR, the N—H combination band (amide bond) at 2055 nm, found in native RNase A, was shifted to 2050 nm upon thermal unfolding. H—D exchange experiments were also used to estimate the number of unexchanged amide protons after exposure to D₂O. The conclusion was that NIR region may be used as a conformation-sensitive monitor of the thermally induced unfolding of proteins in H₂O solutions.

Modeling was performed on polypeptides by Canadian scientists [78] working on medical diagnostic methodology. In their study, NIR PA spectroscopy was used to study 19 homo-polypeptides. The biochemical information was compared with data from MIR. Specific modes were assigned (e.g., the CH-stretch combination region).

Cell bioreactors may be monitored by NIR. Mark Riley et al. [79] reported on fermentation control. The workers used a computer simulation to generate spectra of mixtures of components found in fermentation mixtures. This model allowed for reasonable assay values for a simple binary solution of glucose and glutamine. They then modeled a complex solution containing varying concentrations of ammonia, glucose, glutamate, glutamine, and lactate.

Another fermentation study was presented by Hall et al. in 1996 [80]. A simultaneous NIR assay for acetate, ammonia, biomass, and glycerol was developed for an industrial *Escherichia coli* (*E. coli*) fermentation broth. The PLS equation gave standard errors of 0.7, 1.4, 0.7 g/l, and 7 mmol/l respectively.

33.3.2 (BIO) CHEMISTRY (*IN VIVO*)

One example of this work was performed by Zhang et al. [81]. NIR was correlated with standard pH measurements to perform *in vivo* determination of the myocardial pH during regional ischemia.

Some work was reported on the World Wide Web [82]. Researchers have been working on developing fluorescent probes where markers covalently bind to biomolecules. The penetrating power of NIR light is the driving force behind this research.

33.3.3 CANCER

PAP smears were screened using NIR spectroscopy at Johns Hopkins University [83]. Healthy patients, patients with abnormal cells, and patients with cervical cancer were screened using NIR. Using discriminant analysis PCA the samples were grouped and used to examine other samples. Malignant and healthy tissues were distinctly different, while abnormal tissues carried spectral features from both sets.

Mammograms are uncomfortable and embarrassing for women. NIR imaging [84] and an “optical biopsy” [85] may be performed. NIR radiation has been suggested as an alternative to both x-rays and physical, invasive biopsies. Breast cancer is one of the leading causes of death and disfigurement in woman, it is best detected early [85].

Magnetic resonance imaging (MRI) is used where x-rays are questionable. Using NIR simultaneously could give a better picture of the mass’s chemistry [86]. A time-resolved imager capable of acquiring images simultaneously was used for this work in the range between 780 and 830 nm.

Work was performed by Ntziachristos et al. [86]. In it, they used both MRI and NIR for precise coregistration of images to examine the potential and limitations of optical mammography. Using a time-resolved imager, the group acquired NIR images simultaneous with MR images. The contrast at 780 and 830 nm was used to study the relative enhancement and kinetics due to the administration of Infracyanamine R25, a NIR contrast agent.

In 1994, Meurens et al. [87] was able to determine that cryostat sections of carcinomatous tissue were different spectrally from non-carcinomatous tissue. Four wavelength regions from 1200 to 2370 nm were found best for classification.

33.4 PHYSICS, PHYSICAL PARAMETERS, MATH, AND IMAGING

The advancement of the instrumentation and software is sometimes ignored. A paper presented by Abbot et al. [88] used laser Doppler perfusion imaging to follow skin blood flow. The work was done in red and short wave NIR regions of the spectrum. A multichannel instrument for tissue imaging was developed at the University of Illinois (Urbana-Champaign) [89]. They developed a frequency-domain instrument for noninvasive, real-time, NIR optical tomography of tissue *in vivo*. They also constructed a spatial map of the optical properties of a strongly scattering medium in a semi-infinite-geometry sampling configuration.

In an article by Piantadosi et al. [90], algorithms are the topic of discussion. Using a half-dozen research papers as examples, the authors discuss approaches used in NIR, citing both the hardware and the software.

One paper [91] discussed a technique called fuzzy optimal associative memory (FOAM), used for background prediction of NIR spectra. This software yields better background scans for calculation of NIR spectra of glucose in plasma matrices (from single-beam data).

Arridge et al. [92] published an article on finite element approach for modeling photon transport in tissue. In this method, called finite element modeling (FEM), the photon density inside an object and photon flux at its boundary is introduced into modeling light transport through tissue. They derive a model for one particular case. The calculation of the boundary flux is a function of time resulting from a δ -function point input to a two-dimensional circle (showing as a line source in an infinite cylinder) with homogeneous scattering and absorption properties.

REFERENCES

1. F. M. Ham and G. M. Cohen, Patent, Patent No. 5,553,616 (1997).
2. I. Gabriely et al., *59th Scientific Session, American Diabetes Association Meeting*, San Diego, CA, 1999.
3. M. A. Arnold and L. A. Marquardt, *Anal. Chem.*, 65: 3271 (1993).

4. M. A. Arnold et al., *Appl. Spectrosc.*, 484: 477 (1994).
5. M. A. Arnold et al., *Anal. Chem.*, 687: 1124 (1996).
6. G. W. Small et al., *Appl. Spectrosc.*, 534: 402 (1999).
7. J. J. Burmeister, M. A. Arnold, and G. W. Small, *IEEE Lasers and Electro-Optics Society*, 12: 6 (1998).
8. M. R. Riley, M. A. Arnold, and D. W. Murhammer, *Appl. Spectrosc.*, 5210: 1339 (1998).
9. D. M. Haaland et al., *Appl. Spectrosc.*, 4610: 1575 (1992).
10. W. F. Schrader, "Non-Invasive Anterior Chamber Glucose Monitoring by Near-Infrared Absorption Spectroscopy, an Alternative to Blood-Glucose Monitoring in Diabetic Patients?," *Proc. 96th Determination of Glucose Annual Meeting*, 1998.
11. Patent, J. Backhaus et al., U.S. patent #5,535,743 (July 16, 1996).
12. U. Fischer et al., *Horm. Metab. Res.*, 26: 515 (1994).
13. F. Sternberg et al., *Diabetes Care*, 18: 1266 (1995).
14. M. A. Arnold et al., *Photochem. Photobiol.*, 671: 50 (1998).
15. M. A. Arnold et al., *Analyt. Chem.*, 70: 1773 (1998).
16. J. S. Maier et al., *Optics Lett.*, 1924: 2026 (1994).
17. K. Jagemann et al., *Z. Phys. Chem.*, 191: 179 (1995).
18. C. Fischbacher et al., *Fresenius J. Anal. Chem.*, 359: 78 (1997).
19. K. Danzer et al., *IEEE LEOS Newsletter*, 23–25 (1998).
20. H. Shamoon et al., Amer. Diabetes Assn., Paper 426, San Diego, June 1999.
21. H. M. Heise et al., *Appl. Spectrosc.*, 477: 875 (1993).
22. F. F. Jobsis, *Science*, 198: 1264 (1977).
23. Y. Ozaki et al., *Appl. Spectrosc.*, 461: 180 (1992).
24. M. G. Sowa et al., *Appl. Spectrosc.*, 512: 143 (1997).
25. D. M. Mancini et al., *J. Appl. Physiol.*, 776: 2740 (1994).
26. L. Lin et al., *Proc. of SPIE*, 3257 (Photonics West), Paper 41, San Francisco, January 1998.
27. K. Yamamoto et al., *Proc. of SPIE*, 3257 (Photonics West), Paper 17, San Francisco, January 1998.
28. M. A. Franceschini et al., *Proc. Optical Society of America*, Washington, DC, 1996.
29. K. Feldscher, *The Northeastern Voice*, www.voice.neu.edu/970123/oxygen.html, July 12, 1999.
30. Z. X. Jiang et al., *Proc. of SPIE*, 3257 (Photonics West), Paper 44, San Francisco, January 1998.
31. R. Marbach and H. M. Heise, *Appl. Optics*, 344: 610 (1995).
32. K. Miyasaka, *96 PICU Conference*, Toronto, 1996.
33. S. Van Huffel et al., Department of Pediatrics and Neonatal Medicine, University Hospital Gasthuisberg, Leuven, Belgium, www.esat.kuleuven.ac.be/sista/yearreport96/node6.html
34. S. Van Huffel et al., eee.esat.kuleuven.ac.be/sista/yearreport/node33.html, 1998.
35. C. E. Cooper et al., *Biochem. J.*, 332: 627 (1998).
36. J. S. Wyatt et al., *J. Appl. Physiol.*, 68: 1086 (1990).
37. V. V. Kupriyanov et al., *J. Mol. Cell. Cardiol.*, 29: 2431 (1997).
38. J. R. Mansfield et al., *Computerized Med. Imaging and Graphics*, 215: 299 (1997).
39. T. Wolf et al., *J. Cereb. Blood Flow Metab.*, 16: 1100–1107 (1996).
40. K. D. Liem et al., *European J. Pediat.*, 1537: 504 (1994).
41. H. Nguyen and G. Murphy, Centre for Biomedical Technology, University of Technology, Sydney, Australia, and P. Cooper et al., CRC for Cardiac Technology, July 12, 1999, www.eng.uts.edu.au/~htn/research.html
42. T. J. Germon et al., *J. Clin. Monit.*, 10: 1 (1998).
43. L. C. Henson et al., *Anesthesiology*, 881: 58 (1998).
44. Y. Hoshi et al., *Neurosci. Lett.*, 172: 129 (1994).
45. F. Okada et al., *Euro. Arch. Psychiatry Clin. Neurosci.*, 244: 17 (1994).
46. Y. Hoshi et al., *Brain Res.*, 652: 257 (1994).
47. C. Hirth et al., www.ukrv.de/ch/neuro/hirth.html, 1998.
48. Chow et al., *J. Thorac. Cardiovasc. Surg.*, 1144: 1123 (1997).
49. R. Totaro et al., *Clin. Sci.*, 95: 497 (1998).
50. (News Release), <http://koigakubo.hitachi.co.jp/research/med/release/br.html>
51. E. Dreassi et al., *Analyst*, 1228: 767 (1997).
52. E. Dreassi et al., *Analyst*, 1228: 771 (1997).
53. P. Corti et al., "Application of Near-Infrared Reflectance Spectroscopy in the Study of Atopy. Part 3. Interactions Between the Skin and Fomblins," *Analyst*, 1228: 788 (1997).

54. M. F. Stranc et al., *Brit. J. Plastic Surg.*, 51: 210 (1998).
55. F. Lind, Department of Surgical Sciences, Karolinska Institute, Stockholm, Sweden, <http://research.kib.ki.se/e-uven/public/K3794.html>
56. M. G. Sowa et al., *Vibrat. Spectrosc.*, 10: 49 (1995).
57. N. Kasa and K. M. Heinonen, *Acta Paediatr.*, 82: 1 (1993).
58. K. Z. Liu et al., *Int. J. Gynecol. Obstet.*, 57: 161 (1997).
59. C. H. Barlow et al., in *Optical Tomography, Photon Migration, and Spectroscopy of Tissue and Model Media*, B. Chance and R. R. Alfano (eds.), *SPIE*, 2389: 818 (1995).
60. C. Tsai et al., *Proc. of SPIE*, 3257 (Photonics West), Paper 14, San Francisco, January 1998.
61. J. M. Schmitt et al., *Proc. of SPIE*, 3264, San Francisco, 1998.
62. R. P. Rava et al., *Appl. Spectrosc.*, 462: 187 (1992).
63. S. Anderson-Engels et al., www-lmlc.fysik.lth.se/Prog9395/p43.htm
64. G. Muller et al. (eds.), *Medical Optical Tomography: Functional Imaging and Monitoring*, *SPIE Institute Series*, 11: 397–424, Bellingham, WA, 1993.
65. O. Jariman et al., *Acta Radiolog.*, 33: 277 (1992).
66. R. Berg et al., *Appl. Opt.*, 32: 574 (1993).
67. S. Anderson-Engels et al., *SPIE*, 2081: 137–146, Budapest, Hungary, 1993.
68. R. A. Lodder and L. Cassis, *Spectroscopy*, 57: 12 (1990).
69. *Amer. College Cardiol.*, 48th Annual Sci. Session, New Orleans, March 1999.
70. J. M. Carney, W. Landrun, L. Mayes, Y. Zou and R. A. Lodder, *Anal. Chem.*, 1893, 65: 1305–1313 (1993). Available at http://asrg.contactincontext.org/ASRG/pdfs/gerbil_ocr.PDF. P. R. Moreno, R. A. Lodder, K. R. Purashothaman, W. E. Charash, W. N. O'Connor, J. E. Muller, *Circulation*, 105: 923–927 (2002). Available at <http://asrg.contactincontext.org/ASRG/pdfs/vpcirc.pdf>. R.A. Lodder, Near-infrared Spectrometric Imaging in Stroke Research, available at <http://kerauac.pharm.uky.edu/ASRG/Pittcon/RAL1995/ROBPIT.html>
71. R. J. Dempsey, L. A. Cassis, D. G. Davis, R. A. Lodder, Near-infrared imaging and spectroscopy in stroke research: Lipoprotein distribution and disease, available at <http://www.pharm.uky.edu/ASRG/Wave/Lipo/lipo.htm>
72. M. J. McShane and G. L. Cote' *Appl. Spectrosc.*, 528: 1073 (1998).
73. J. Hall, *Proc. of SPIE*, 3257 (Photonics West), San Francisco, January 1998.
74. R. A. Shaw et al., "Quantitation of Protein, Creatinine, and Urea in Urine by Near-Infrared Spectroscopy," *Clin. Biochem.*, 291: 11 (1996).
75. M. Jackson et al., *Biophys. Chem.*, 68: 109 (1997).
76. R. A. Shaw et al., *Reumatol. Int.*, 15: 159 (1995).
77. C. P. Schultz et al., *Biospectroscopy*, 4: S19 (1998).
78. J. Wang et al., *J. SPIE*, 2089: 492 (1996).
79. M. R. Riley et al., *AICHE 1999 Annual Meeting*, Session 280–non-invasive measurements, Dallas, Texas.
80. J. W. Hall et al., *Appl. Spectrosc.*, 501: 26 (1996).
81. S. Zhang, *Proc. of SPIE*, 3257 (Photonics West), Paper 13, San Francisco, January 1998.
82. Regensburg University, <http://pc3898.uni-regensburg.de/Wolfbeis/et/labels.html>, Institute of Anal. Chem., 1996.
83. Zhengfang Ge et al., *Appl. Spec.*, 494: 1324 (1995).
84. V. Ntziachristos et al., University of Pennsylvania, www.lrsmpenn.edu/~vasilis/Concurrent.html
85. University of Illinois Urbana-Champaign, *ScienceDaily*, July 13, 1999, www.sciencedaily.com
86. V. Ntziachristos et al., www.lrsmpenn.edu/~vasilis/frresearch.html, University of Pennsylvania, 1997.
87. M. Meurens et al., *Appl. Spectrosc.*, 482: 190 (1994).
88. N. C. Abbot et al., *J. Investigative Dermatology*, 1076: 2235 (1996).
89. M. A. Franceschini et al., presented at the 1995 SPIE Conference, on Web page of the University of Illinois at Urbana-Champaign, www.physics.uiuc.edu/groups/fluorescence/spie95
90. C. A. Piantadosi et al., *Analyt. Biochem.*, 253: 277 (1997).
91. P. B. Harrington and B. W. Wabuyele, *Appl. Spectrosc.*, 501: 34 (1996).
92. S. R. Arridge et al., *Med. Phys.*, 202: 299 (1993).

34 Near-Infrared Spectrometry in Cardiovascular Disease

Aaron A. Urbas and Robert A. Lodder

CONTENTS

Acknowledgments	670
References	670

Near-infrared (NIR) spectrometry can be a useful tool in the study of cardiovascular diseases (CVDs) including atherosclerosis and aneurysm. Atherosclerosis in the vessel wall has been linked to aneurysm, although it is not clear that this is a causal relationship. Studies have been undertaken in both animal models of CVD and in human patients. There is widespread agreement that new diagnostic techniques are required to identify coronary plaques that are prone to disruption. The type of plaque considered most vulnerable to disruption is a thin-capped fibroatheroma with increased inflammatory cell content. Multiple techniques are being tested to identify such plaques before they disrupt and cause thrombosis. Identification of these potentially lethal plaques before they disrupt will facilitate the development of therapeutic strategies to prevent acute coronary events. In animal research, several genetic “knockout” mouse models have been developed recently to mimic human atherosclerosis [1] and abdominal aortic aneurysm (AAA) [2]. Techniques for monitoring the onset, progression, and regression of these processes in mouse models could provide valuable pathophysiological insights into the disease processes. Nondestructive *in vivo* techniques will be needed for proteomics studies in these models. Finally, these analytical methods may be useful in assessing the effectiveness of possible treatments. Atherosclerosis is a chronic inflammatory process [3–5] with complications that are the leading cause of death in Western societies. Extensive research has been done to determine the complex pathophysiology of atherosclerosis, although mechanisms for various aspects are still being elucidated. Among the earliest changes in the vessel wall is an increase in retained lipoproteins [6–8] and subsequent oxidation [7,9,10] in the subendothelial matrix. Development of lipid-laden macrophages (foam cells) is another hallmark of the early atherosclerotic process [11,12]. Proliferation and phenotypic changes in smooth muscle cells are seen as well [13,14]. The advanced atherosclerotic lesion may be characterized by accumulation of extracellular lipid, development of a lipid-rich necrotic core, formation of a fibrous cap, and calcification [15].

Abdominal aortic aneurysms represent potentially life-threatening conditions that occur in up to 10% of the aged populations in industrialized nations. An aneurysm is broadly defined as a permanent localized dilatation of an artery. AAAs arise because of substantial remodeling of the extracellular matrix and are frequently accompanied by atherosclerosis. They may be manifested by catastrophic rupture, signs of pressure on other viscera, or an embolism originating in the aneurysm wall, but most are asymptomatic.

Collagen and elastin are major structural components of vessel walls that have been widely implicated in aneurysm formation, progression, and rupture. The most prevalent structural modification associated with human AAAs that has been reported is a reduction in elastin concentration in the aortic wall [16–22]. Significant correlations between reduced elastin concentration and AAA diameter have been observed [19]. Alternatively, other studies have shown that reduction in elastin concentration is essentially complete before dilation of AAAs [21,23]. One proposed mechanism for reduced elastin concentrations is degradation or loss brought about by elastolysis [16–18,21]. Other work has reported that elastin content in the vessel walls of AAAs actually increases [22,24]. In these studies, a 2.5-fold increase in elastin content was found in AAAs vs. normal aortic samples of equal length. This increase, however, was accompanied by a significantly greater increase in total matrix proteins, which suggests that reduction in elastin content is at least in part due to dilution. These results and work by others suggest that an important mechanism in AAA formation is the regulation of matrix macromolecule synthesis [22,24,25].

Increased collagen concentration is another matrix modification that has been widely observed in human AAAs [18–20,22,25]. Modifications in collagen organization and deposition have been correlated to rupture in human AAAs [17,19,21,26]. Although there are noticeable differences in the findings of these studies, it is evident that increases in the collagen-to-elastin ratio are a general observation in AAAs.

Diffuse reflection NIR spectroscopy has proven to be a useful technique for identifying chemical content of biological tissues [27]. Biological applications of NIR spectroscopy include monitoring systemic and cerebral oxygenation and identifying plasma constituents including glucose, total protein, triglycerides, cholesterol, urea, creatinine, and uric acid [28–32]. Our group has reported on the use of NIR spectroscopy to classify human aortic atherosclerotic plaques and to identify cholesterol, high-density lipoprotein (HDL), and low-density lipoprotein (LDL) in arterial wall samples [33–35].

Preliminary research has been conducted to determine the feasibility of using NIR spectroscopic methods to monitor collagen and elastin in the aortas of ApoE knockout mice with atherosclerosis and aneurysm formed from chronic infusion of the peptide angiotensin II (Ang II). The goal of such research is to eventually create a novel fiber-optic probe capable of collecting NIR spectra in mouse aortas *in vivo*.

The mouse model has been detailed elsewhere [36], and only a brief description will be presented here. Infusion of Ang II into mature apolipoprotein E-deficient ($\text{apoE}^{-/-}$) mice promotes an increase in the severity of aortic atherosclerotic lesions and the formation of AAAs. The formation of aneurysms in the animal model is independent of arterial blood pressure and lipoprotein profiles; however, it requires the hyperlipidemic background and is dependent on Ang II dose and gender (males develop aneurysms at a greater incidence than females).

Female $\text{apoE}^{-/-}$ mice (back-crossed 10× into the C57BL/6J background) have been implanted with Alzet osmotic minipumps when the mice were at 6 months of age. Pumps were filled either with saline vehicle or with solutions of Ang II that delivered (subcutaneously) either 500 (medium-dose group) or 1000 (high-dose group) ng/kg/min Ang II for 28 days. At this point, the aortas were removed and analyzed. In these mice, aortic tissue was removed from the ascending aorta to the ileal bifurcation and placed in 4% paraformaldehyde in phosphate-buffered saline (PBS) overnight at room temperature. The intimal surface was exposed by a longitudinal cut through the inner curvature down the whole length of the aortic tree. Each aortic sample was cut into six roughly equal sections, resulting in 36 total samples.

Reference spectra of collagens I, II, III, and IV and elastin were obtained using a scanning monochromator system (see Figure 34.1 and Figure 34.2). Collagens I and III are the principal collagen components of the aorta. NIR spectra of the aorta samples were collected with a modified filter photometer system sometimes used as a high-intensity light source for NIR arterial catheters. Each sample was placed on an ordinary microscope slide for analysis. The samples were not immersed in solution during spectra collection. The instrumental configuration allowed the entire tissue sample to contribute to the NIR signal. NIR reflectance measurements were made at 20 wavelengths between

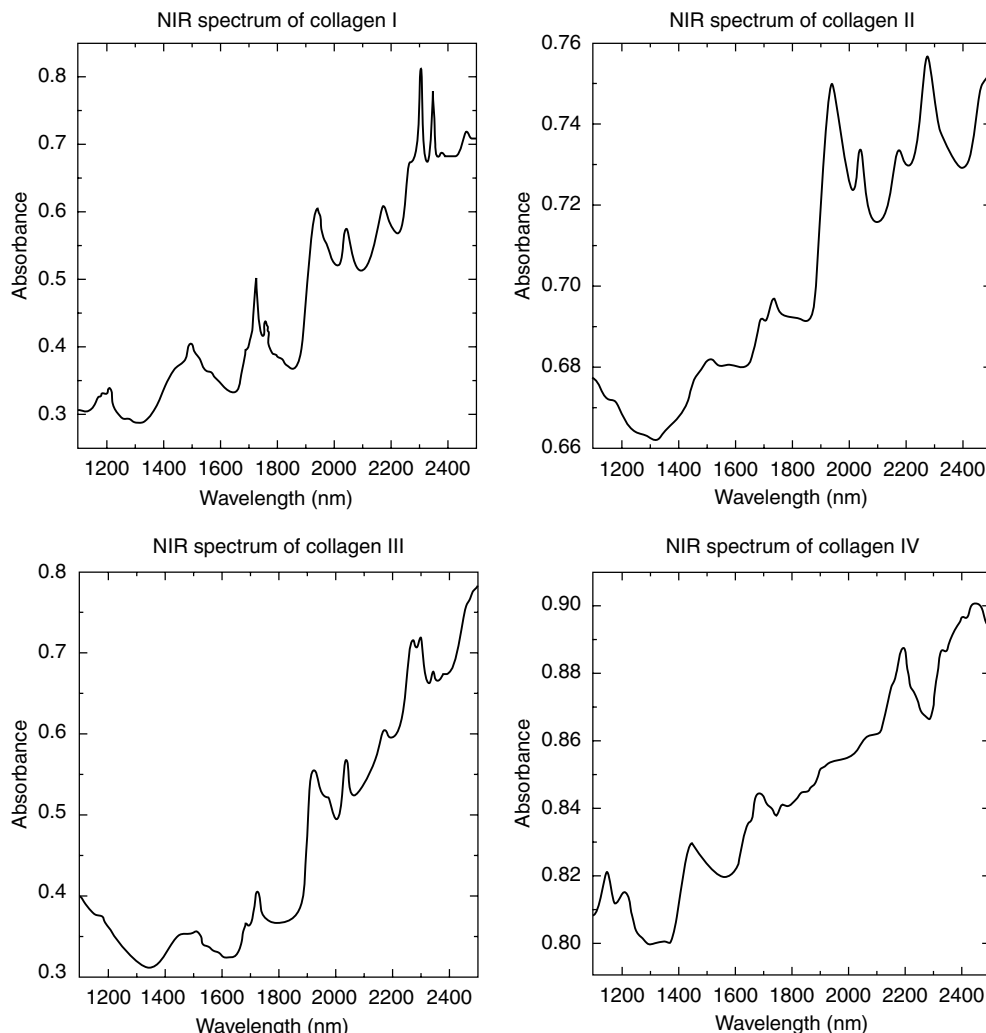


FIGURE 34.1 NIR spectra of collagen standards. Collagens I and III are the principal collagens of the aorta. The spectra of the four collagens are distinctive, suggesting that simultaneous multicomponent analysis of collagens is possible.

1445 and 2350 nm. A cubic spline-fitting algorithm obtained continuous spectra from the absorbance values. The spectral data were scatter-corrected before data analysis. Spectra from a control and a high-dose tissue sample are presented in Figure 34.3 for comparison. The chemical composition of the tissue samples between groups is similar, and as a result, the gross appearances of the two spectra are similar. Collagen-to-elastin ratios for tissue samples were obtained by scanning electron microscopy (SEM) and morphometry. Detailed information on this technique can be found in the literature [37,38]. To determine collagen-to-elastin ratios, three SEM images were obtained for each sample corresponding to the inner and outer membranes and a cross section of the aortic wall.

Principal component analysis (PCA) and principal component regression (PCR) were used to analyze the data [39,40]. PCR was used to construct calibration models to predict Ang II dose from spectra of the aortas. A cross-validation routine was used with NIR spectra to assess the statistical significance of the prediction of Ang II dose and collagen/elastin in mice aortas. The accuracy of the PCR method in predicting Ang II dose from NIR spectra was determined by the *F* test and the standard error of performance (SEP) calculated from the validation samples.

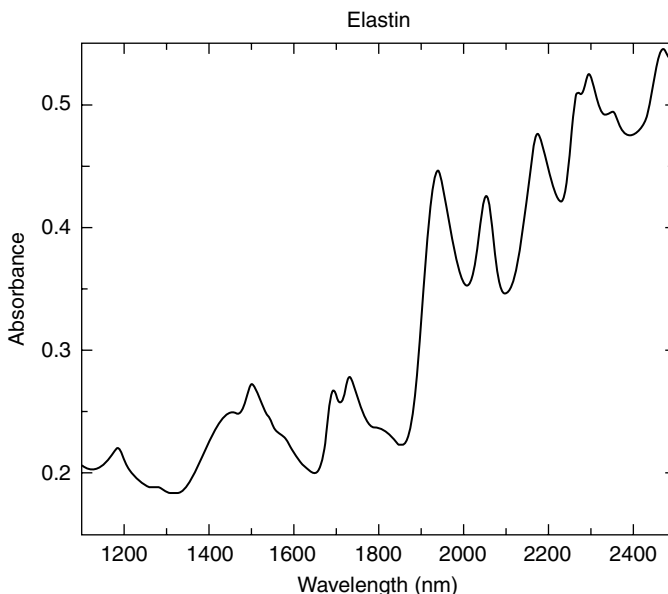


FIGURE 34.2 NIR spectrum of elastin standard. Elastin has a unique spectrum that differentiates it from the collagens.

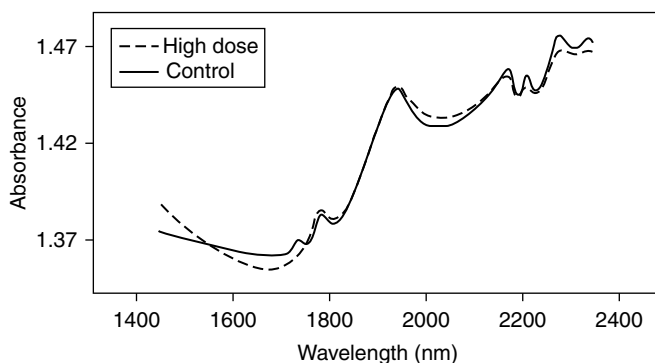


FIGURE 34.3 Examples of the NIR spectra obtained from aorta sections of a high Ang II dose (1000 ng/kg/min) mouse and control mouse.

The spectra were investigated first by PCA to examine the relationships between the samples. The first two PCs accounted for approximately 89% of the variation in the data set (see Figure 34.4) [41]. The spectral separation among the different groups (related to Ang II dose) is significant. Also evident is the comparatively minor variation between spectra within each group (the scatter in each group comes from different mice and the position of the aortic section scanned). The control and high-dose groups contain two and three mice aortas, respectively. The medium-dose group contained samples from a single mouse.

The ability to predict Ang II dose from the NIR spectra was examined using PCR. The SEP provides a global estimate of the prediction capabilities of the NIR-PCR method. The SEP for Ang II dose depends on the number of PCs used in the model. The first two PCs provided the optimum prediction models, and the SEP for Ang II dose by PCR was 37 ng/kg/min (standard error of estimate [SEE] = 26 ng/kg/min, $r^2 = .99$, $f_{0.01}$ significance).

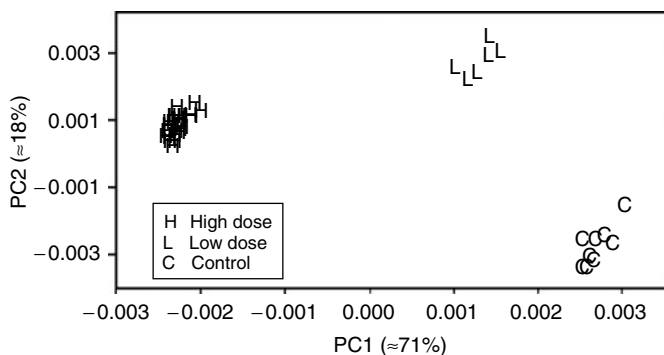


FIGURE 34.4 First two PCs of the NIR spectra of the aorta data set. H denotes the spectra from aortic sections of the high-dose mice (1000 ng/kg/min Ang II, $n = 18$ samples), L denotes the spectra from aortic sections of the low-dose mice (500 ng/kg/min Ang II, $n = 6$ samples), and C denotes the spectra from aortic sections of the control mice (saline, $n = 12$ samples).

Collagen-to-elastin ratios in the samples were estimated using SEM to determine whether a correlation existed between the NIR spectra and this histological marker. The range of collagen-to-elastin ratios obtained was 1.4 to 4.5. A trend apparent in the data is a general increase in collagen-to-elastin ratio with increasing Ang II infusion (SEE = 0.38, SEP = 0.39, $r^2 = .85$, $f_{0.05}$ significance). Comparison of these data with the NIR spectra illustrates that the within-group variation is greater in the collagen-to-elastin ratios than in the NIR spectra. The $r^2 = .84$ between the collagen-to-elastin ratio determined by SEM morphometry and the actual Ang II dose. Two factors that likely contribute to the error in collagen-to-elastin ratio are the limited scope of the SEM image sampling and the fact that collagen and elastin are not the only components contributing to the NIR spectra. The entire aortic sample contributes to the NIR spectra. Substantial lipid signals are observed in the spectra as well as signals from collagen and elastin. SEM sampling is limited to three SEM images from the inner and outer membranes and a single cross section of each vessel segment.

This work presents a macroscopic study of arterial changes in mice aorta by NIR spectroscopy. Figure 34.1 demonstrates that collagens and elastin have very distinctive NIR spectra. Collagens I and III are the principal collagens of the aorta. Figure 34.4 suggests that diffuse alterations occur in the aortic walls of Ang II-infused apoE^{-/-} mice and that these changes can be observed with NIR spectroscopic analysis of intact tissue. The better correlation between the NIR spectra and the Ang II dose than between the NIR spectra and the collagen-to-elastin ratio suggests that there is more occurring in the aortas than collagen/elastin changes and that the NIR spectra detect these additional chemical changes. This suggestion is not surprising given that NIR spectra show contributions at some level from virtually every organic compound in any tissue.

Though the data are limited, they also indicate that the NIR identifiable characteristics of the aorta are consistent between mice within the same treatment group (this cannot be established with respect to the medium-dose group, however, because the tissue samples came from a single mouse). Together, the results support the development of a catheter and extended-use fiber-optic implant for use in the blood vessels of mice and suggest wavelengths that might be important to monitor.

In an effort to determine which spectral changes in the aortas were associated with collagens I and III and elastin composition changes, a set of sample mixtures of collagens I and III and elastin was prepared using pure lyophilized standards. Figure 34.5 shows the composition of each of the prepared sample standards, with the pure collagen I (C1) standard in one corner of the triangle, the pure collagen III (C3) in another corner of the triangle, and finally the pure elastin in the remaining corner of the triangle. The concentrations of each constituent in the standard mixtures were set at 0, 25, 50, 75, or 100 wt.% of each lyophilized protein. The vertexes represent all possible combinations of mixtures in the percentages given (a total of 15 mixtures including the pure corner standards).

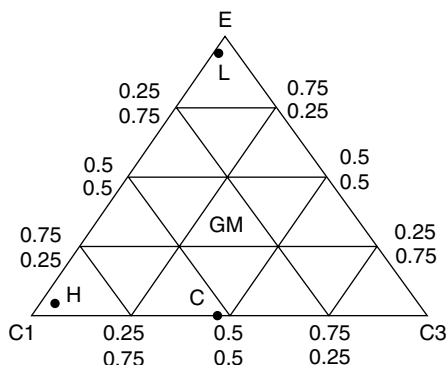


FIGURE 34.5 Diagram showing the composition of each of the prepared sample standards. The pure collagen I (C1) standard is in one corner of the triangle, the pure collagen III (C3) is in another corner of the triangle, and the pure elastin (E) in the remaining corner of the triangle. The concentrations of each of the three constituents in the standard mixtures were set at 0, 25, 50, 75, or 100 wt.% of each lyophilized protein. The vertexes in the triangle represent all possible combinations of protein mixtures in the percentages given (a total of 15 mixtures including the pure corner standards). GM, the group mean, is the center of the triangle and represents a mixture of one third of each protein.

The center (i.e., group mean, or GM in Figure 34.5) would represent a mixture of one third of each protein, but this sample was not actually prepared in the set.

The reflection spectra of the 15 mixtures were compared with the reflection spectra of the 36 aorta sections by mean-centering the spectra of the mixtures and the spectra of the aortas. The difference spectra between each standard sample spectrum and the mean spectrum of the standard samples were calculated. Likewise, the difference spectra between each aorta section spectrum and the mean spectrum of the aorta sections were also calculated. The aorta difference spectra were then averaged for the control, low-dose, and high-dose Ang II groups. Finally, the difference spectra of the standards and the averaged aorta sections were then correlated using $r = \Sigma(s_1 s_2) / (\Sigma s_1^2 \Sigma s_2^2)^{1/2}$, where s_1 is spectrum 1 from the set of standards and s_2 is spectrum 2 from the set of aorta sections [41].

The correlations between the average control group (saline-infused) spectrum (0 ng/kg/min Ang II) and each standard sample spectrum ranged between ± 0.97 . The highest correlation was with the (50, 50, 0%) (C1, C3, elastin) standard. The correlations between the average low-Ang II dose group spectrum and each standard sample spectrum ranged between ± 0.99 . The highest correlation of the average spectrum of the low-Ang II dose group was with the pure elastin (0, 0, 100%) standard. The correlations between the average high-dose group spectrum and each standard sample spectrum ranged between ± 0.87 . The highest correlation was with the C1 (100, 0, 0%) standard. These high correlations suggest that much (but not all) of the variations in aorta spectra with Ang II dose can be attributed to changes in collagen and elastin composition. The locations of the mean control (C), low-dose (L), and high-dose (H) Ang II aortic sections are shown in Figure 34.5. The H, L, and C points are interpolated to the location of maximum correlation to the standards. The spectra of the most highly correlated standard samples and each treatment group is shown in Figure 34.6. The correlation between the aortic sections and the standard mixtures is the highest for the control and low-dose Ang II groups in Figure 34.6. The observation that the changes in the spectra of the standard mixtures do not add up perfectly to the changes found in the aortas also suggests that there is more occurring in the aortas than collagen/elastin changes and that the NIR spectra detect these additional chemical changes. The majority of these additional chemical changes appear to be in the C—H stretching overtone regions of the spectra (the lipid region; see Figure 34.3). Figure 34.3 shows a bathochromic shift in the C—H stretching overtones that usually corresponds to an increase in trans or saturated lipids [35].

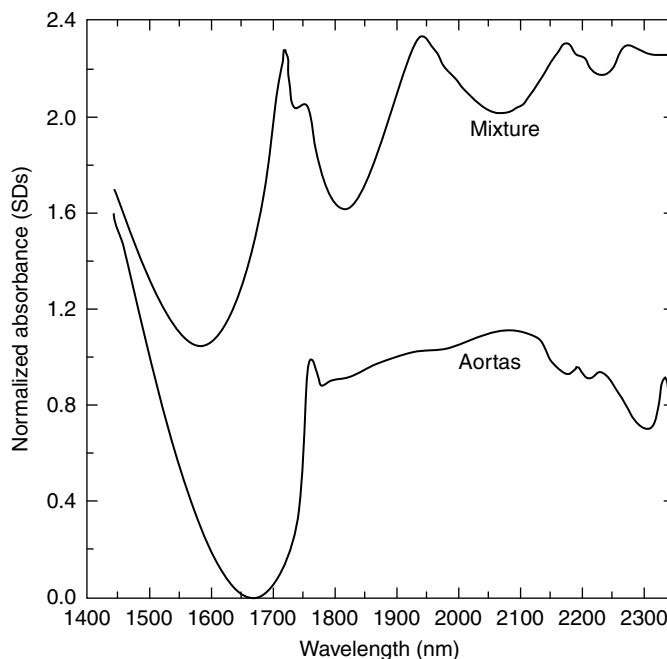


FIGURE 34.6 Spectra of the most highly correlated standard mixture of collagens and elastin and mean aorta spectrum for each treatment group: (top left) control group (saline infusion); (top right) low Ang II dose, 500 ng/kg/min Ang II; (bottom) high Ang II dose, 1000 ng/kg/min. The spectra in each treatment group are distinctive. The correlation between the aortic sections and the standard mixtures is the highest for the control and low-dose Ang II groups.

This preliminary study of aneurysm has some limitations. The small number of mice (six) used as the source of the 36 aorta sections limits the observable variation in the data set. The lack of detailed histological data for the samples prevents the association of the spectra with specific tissue pathologies and comparison of pathology between samples. Studies with larger, more diverse populations are necessary to better investigate the spectral trends and variation between mice.

However, NIR spectra are distinctive for proteins in the blood vessel wall (specifically collagens and elastin). The ability of NIR spectrometry to collect useful spectra in aqueous environments may make it useful for proteomics *in vivo*. The results of this study suggest that NIR spectroscopy is a potentially useful technique for investigating vascular changes and protein composition associated with AAA in a mouse model of this disease. These results support an expanded study in the future to correlate NIR spectra with chemical compositions and histological features in aortic aneurysm, which is known to often be associated with atherosclerosis.

For studies of atherosclerosis, a method is needed to identify nonstenotic, lipid-rich coronary plaques that are likely to cause acute coronary events. NIR spectroscopy can provide information on the chemical composition of tissue. NIR spectroscopy can identify plaque composition and features associated with plaque vulnerability in human aortic atherosclerotic plaques obtained at the time of autopsy. In one study, 199 samples from 5 human aortic specimens were analyzed by NIR spectroscopy [35]. Features of plaque vulnerability were defined by histology as presence of lipid pool, thin fibrous cap (65 μm by ocular micrometry), and inflammatory cell infiltration. NIR spectroscopy sensitivity and specificity for histological features of plaque vulnerability were 90 and 93% for lipid pool, 77 and 93% for thin cap, and 84 and 89% for inflammatory cells, respectively. NIR spectroscopy can identify plaque composition and features associated with plaque vulnerability in postmortem human aortic specimens. These results support efforts to develop an NIR spectroscopy catheter system to detect vulnerable coronary plaques in living patients.

Cardiovascular diseases has been the primary cause of death in industrialized countries for some time, and it is rapidly becoming the number one killer in the developing countries [42]. According to recent estimates, 61,800,000 Americans have one or more types of CVD. Each year, more than 1 million people in the United States and more than 19 million others worldwide suffer a sudden cardiac event (acute coronary syndromes and sudden cardiac death). A considerable segment of this population has no preceding symptom. There is a mandate for diagnosis and treatment of the pathological conditions that lie beneath these sudden cardiac events, and identifying vulnerable plaques and patients.

The word “vulnerable” is used to denote the probability of exhibiting a clinically significant event in the future. The word vulnerable has been employed in a variety of reports in the medical literature, all of which portray conditions predisposed to injury. In this respect, the term “vulnerable plaque” is most appropriate to classify plaques susceptible to complications. In contrast, interventional cardiologists and cardiovascular pathologists retrospectively explain the plaque responsible for coronary occlusion and death as a *culprit* plaque, apart from its histopathological appearance. However, for prospective evaluation, diagnosis, and treatment, clinicians require a term like culprit for identifying such plaques *before* an event occurs.

Plaque rupture is the most frequent type of plaque complication, accounting for an excess of 70% of fatal acute myocardial infarctions and sudden coronary deaths. A number of retrospective autopsy series and a handful of cross-sectional clinical studies have indicated that thrombotic coronary death and acute coronary syndromes are instigated by plaque features and associated factors. The majority of methods for detecting and treating vulnerable plaque are dedicated to rupture-prone plaque. This class of plaque is commonly called a “thin-cap fibroatheroma.”

A thin-cap fibroatheroma is typified by a large lipid core rich in cholesterol and cholesterol esters. These plaques have a cap thickness of less than 100 μm and a lipid core accounting for greater than 40% of the plaque’s total volume [43]. Potential *in vivo* intravascular diagnostic techniques include optical coherence tomography (OCT), intravascular ultrasonography (IVUS), elastography (palpography), magnetic resonance imaging (MRI), angiography, and NIR spectroscopy. Increasing evidence substantiates that diverse types of vulnerable plaque with differing histopathology and biology exist. Autopsy [44] and IVUS studies [45] have demonstrated that atherosclerotic lesions commonly exist in young and asymptomatic persons. What percentage of these lesions represents morphologies of rupture-prone vulnerable plaques remains to be determined. Furthermore, chronic inflammation [46] and macrophage/foam cell formation are a fundamental element of the natural history of atherosclerosis. To assess plaque vulnerability, it is apparent that a collective methodology able to appraise structural characteristics (morphology) as well as functional properties (activity) of plaque will likely be most revealing, and may offer higher prognostic value than a single method.

Among the first changes in the arterial wall in atherosclerosis is an increase in retained lipoproteins and ensuing oxidation in the subendothelial matrix [41]. Development of lipid-laden macrophages (foam cells) is another characteristic of the early atherosclerotic progression. Proliferation and phenotypic alterations in smooth muscle cells are also observed. The highly developed atherosclerotic lesion may be distinguished by amassing of extracellular lipid, growth of a lipid-rich necrotic core, establishment of a fibrous cap, and calcification. Atherosclerosis in the arterial wall is also correlated to aneurysm.

Diffuse reflection NIR spectroscopy has proven to be a useful technique for identifying chemical content of atherosclerotic tissues [35]. Our laboratory has described the use of NIR spectroscopy to categorize human aortic atherosclerotic plaques and to quantify cholesterol HDL, and LDL in arterial wall samples [33,34].

Synchrotron infrared (IR) light is approximately 1000 times more intense than a conventional IR source. In addition, synchrotron IR light is highly collimated like a laser, making it more easily focused onto a small spot. However, unlike a laser, the synchrotron emits a wide range of IR wavelengths, enabling Fourier transform infrared (FTIR) microspectroscopy. Consequently, with synchrotron IR light, samples that are smaller and more dilute in concentration can be studied.

In addition, the 1000-fold increase in brightness translates to data collection times that are about 30 times faster with the synchrotron source in comparison to a global source. In studies of excised plaque, synchrotron radiation increases the spatial resolution of collagens and elastin attainable at the site of plaque rupture. Measuring IR fundamentals provides stronger signals with less peak overlap than those available with NIR overtone and combination bands.

Similar composition changes have been observed in both the fibrous cap of lesions in atherosclerosis and in the media and adventitia of AAAs, albeit on different physical scales. A preliminary study tested the hypothesis that the changes observed in C1, C3, and elastin in aneurysm on a millimeter scale are similar to the changes that occur in the fibrous cap of vulnerable atherosclerotic plaque on a scale of micrometers or tens of micrometers. If successfully demonstrated, the similarities could be used as *in vivo* markers of the vulnerable plaques most in need of treatment, and could be used in monitoring therapies in atherosclerotic plaques treated by drugs.

Twenty human coronary tissue sections were obtained postmortem and mounted in paraffin onto low-e glass slides for IR microspectroscopy. A visible light image of a stained section of human coronary atherosclerotic plaque appears in Figure 34.7. This section was adjacent to the section used for IR microspectrometric imaging (see Figure 34.8).



FIGURE 34.7 Stained section of human coronary vulnerable atherosclerotic plaque, obtained as a visible light image. This section was adjacent to the one used for IR microspectrometric imaging.

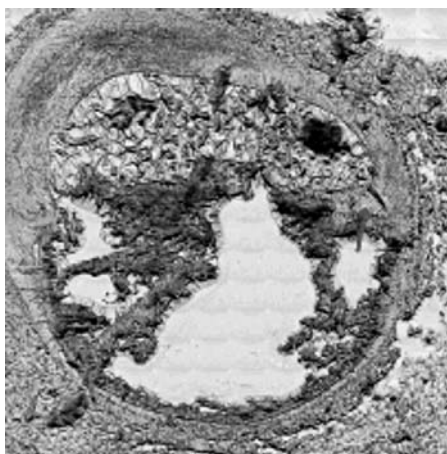


FIGURE 34.8 Unstained section of human coronary vulnerable atherosclerotic plaque, mounted in paraffin. An arrow marks the location on the fibrous cap from which IR spectra were obtained.

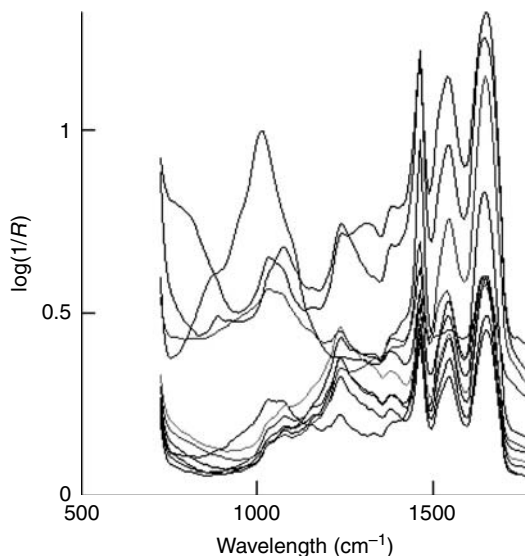


FIGURE 34.9 Representative IR spectra from the region near the arrow in Figure 34.8. The sharp peak at 1469 cm^{-1} arises from paraffin.

An arrow marks the location on the fibrous cap from which IR spectra were obtained. Figure 34.9 depicts representative spectra from a small area of the fibrous cap. A total of 80,600 spectra were obtained from all plaque sections.

The IR microspectrometer at beamline U2b of the vacuum ultraviolet (VUV) storage ring of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL), Upton, NY, was used to collect spectra. Schwartzchild $32\times$ and $10\times$ all reflecting mirror lenses were used for the objective and condenser, respectively. A remote projected image plane mask before the objective produced the apertures used for single point spectra or raster scan mapping via a digitally controlled motorized microscope stage. Spectra were recorded in a reflection absorption mode. A clear location on the IR reflecting microscope slide was used to obtain a reflection-background spectrum and a paraffin-background spectrum.

Mapping was also accomplished utilizing a global source focal plane array instrument. The Perkin-Elmer Spotlight model 300 was employed to obtain rectangular maps of select regions of the sections being examined. For focal plane array images, the $6.25 \times 6.25\text{ }\mu\text{m}$ pixel size was used.

Preliminary examination of each map was done from a locally baseline corrected peak area for the triplet at 1236 cm^{-1} and the doublet associated with the 1082 cm^{-1} band. A map of the ratio of the area of the 1236 cm^{-1} band to that of the 1082 cm^{-1} was used to locate the region with the highest relative amount of the C1.

Reference FTIR spectra of C1, C3, and elastin (Sigma) were obtained (see Figure 34.10).

While such reference compounds are sometimes contaminated by small amounts of lipid, potential contamination posed no problem for this research because the lipid regions of the spectrum were not used in the analysis to avoid the paraffin. The spectral data were scatter-corrected before data analysis (see Figure 34.11).

Representative spectra from a human coronary tissue section are presented in Figure 34.9 for comparison with Figure 34.11. The sharp peak at 1469 cm^{-1} in Figure 34.9 arises from paraffin. The chemical composition of the tissue samples between adjacent areas of tissue is typically similar, and as a result, the gross appearances of the spectra are similar.

In an effort to determine which spectral changes in the coronary sections were associated with C1 and C3 and elastin composition changes, a set of sample mixtures of C1, C3, and elastin was prepared using pure lyophilized standards (see Figure 34.11). A triangular array of compositions

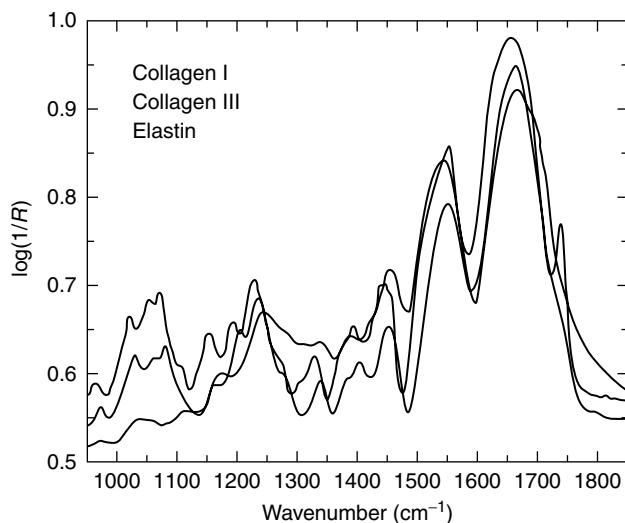


FIGURE 34.10 Spectra of lyophilized standards of collagen I, III, and elastin.

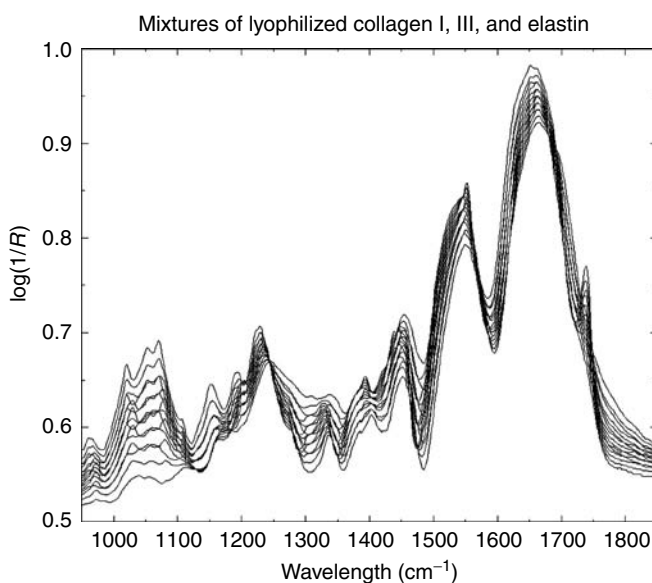


FIGURE 34.11 Spectra of standard mixtures of collagen I, III, and elastin used for mean-centered correlation analysis.

was constructed for this study similar to the one used in Reference 48. The composition of each of the prepared sample standards was represented by a vertex in the array, with the pure C1 standard in one corner of the triangle, the pure C3 in another corner of the triangle, and finally the pure elastin in the remaining corner of the triangle. The concentrations of each constituent in the standard mixtures were set at 0, 25, 50, 75, or 100 wt.% of each lyophilized protein. The vertexes represent all possible combinations of mixtures in the percentages given (a total of 15 mixtures including the pure corner standards). The center (i.e., group mean) would represent a mixture of one third of each protein, but this sample was not actually prepared in the set. The reflection spectra of the 15 mixtures were compared with the reflection spectra of the coronary sections by mean-centering the spectra of the mixtures and the spectra of the coronary artery sections. The difference spectra between

each standard sample spectrum and the mean spectrum of the standard samples were calculated. Likewise, the difference spectra between each coronary section spectrum and the mean spectrum of the coronary sections were also calculated. Finally, the difference spectra of the standards and the coronary sections were then correlated using the product-moment correlation coefficient. The values of these correlation coefficients were contour plotted to produce images corresponding to C1, C3, and elastin in Figure 34.12 to Figure 34.15.

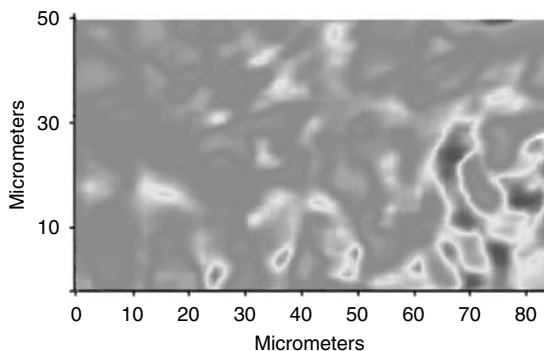


FIGURE 34.12 (See color insert following page 622.) Collagen I distribution in targeted region of fibrous cap (red = high, blue = low concentration).

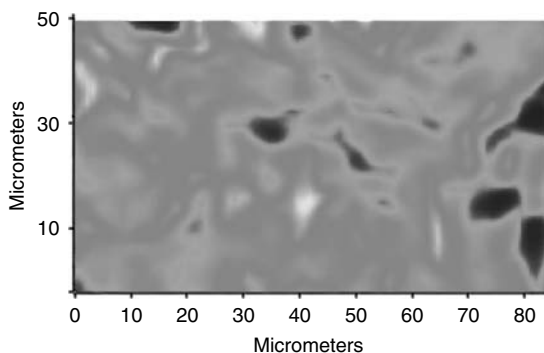


FIGURE 34.13 (See color insert following page 622.) Collagen III distribution in targeted region of fibrous cap (yellow = high, blue = low concentration).

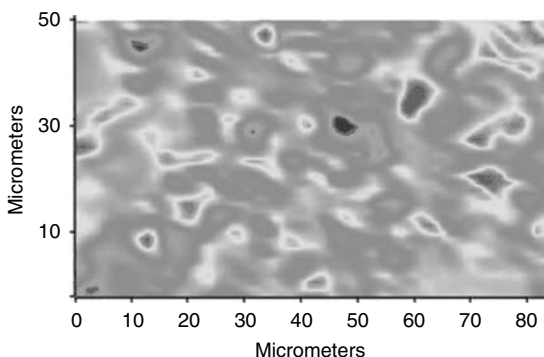


FIGURE 34.14 (See color insert following page 622.) Elastin distribution in targeted region of fibrous cap (red = high, blue = low concentration).

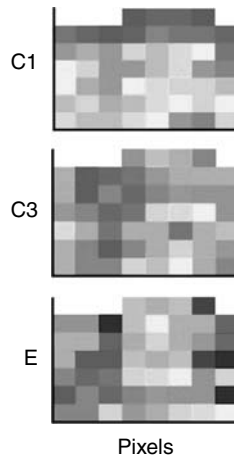


FIGURE 34.15 (See color insert following page 622.) Localized distribution of collagen I, III, and elastin at individual pixel resolution.

The correlations between the coronary tissue section spectra and the set of standard sample spectra ranged between ± 0.99 . In consequence, the contours in Figure 34.12 to Figure 34.14 covered a wide range of correlation values (as did the pixels in Figure 34.15). Color was used to represent correlation between the spectra of the coronary sections and reference standards. Violet represents the lowest correlation, while red represents the highest. The region at the far left of Figure 34.12 to Figure 34.14, off the edge of the thin cap, served as a sort of internal standard or control for the images as the scanning progressed toward the right and the thinnest portion of the cap.

Figure 34.12 shows the correlation between the tissue spectra and amount of C1 in the standards in the region marked by the arrow in Figure 34.8. Pixel values were interpolated between the concentrations of the standards to reach the maximum correlation, as performed in Reference 41. While the amount of C1 does not increase monotonically from left to right in the image, there is a gradient in concentration over the 80- μm distance across the fibrous cap. The right side of the image in Figure 34.6 is the side closest to the thinnest area of the fibrous cap, as shown in Figure 34.7.

Figure 34.13 depicts the distribution of C3 predicted in a similar manner. C3 also shows an overall gradient, but instead generally decreases nonmonotonically from left to right in the image. In contrast, the amount of elastin did not appear to change substantially across the field of the image (see Figure 34.14). However, most of the loss of elastin could have occurred long before the coronary tissue was collected. The control section imaged at the bottom of the vessel was similar to Figure 34.14 in that it showed no obvious concentration trends for either the collagens or elastin. Overall, the trends in C1, C3, and elastin observed in human coronary plaque are similar to those observed in AAA. At single pixel resolution on the right side of the sample zone nearest the thinnest region of the fibrous cap (see Figure 34.15), gradients are not as clear.

Rupture-prone plaques are not the lone vulnerable plaques. All categories of atherosclerotic plaques with high probability of thrombotic complications and swift progression should be regarded as vulnerable plaques. Furthermore, vulnerable plaques are not the only culprit features leading to the occurrence of acute coronary syndromes, myocardial infarction, and sudden cardiac death. Vulnerable blood (i.e., blood inclined toward thrombosis) and vulnerable myocardium (inclined toward lethal arrhythmia) perform an essential role in the clinical outcome. The phrase “vulnerable patient” has even been proposed for the classification of persons with high probability of emergent cardiac events in the near future. A quantitative means of cumulative risk assessment of vulnerable patients should be created that includes variables describing plaque, blood, and myocardial vulnerability. Newly developed assays (e.g., for C-reactive protein), imaging techniques (e.g., computed

tomography [CT] and MRI), noninvasive electrophysiological tests (for vulnerable myocardium), and promising catheters (to localize and characterize vulnerable plaque), together with prospective genomic and proteomic methods, will lead researchers in the hunt for vulnerable patients. These analytical methods will also lead to the expansion and exploitation of new therapies, and eventually to reduction in the incidence of acute coronary syndromes and sudden cardiac death.

This preliminary study of atherosclerosis also has some important limitations. A single patient served as the source of the 80,600 spectra collected from 24 coronary sections, limiting the observable variation in the data set. The lack of detailed histological data for the samples and lack of clinical history from the patient prevent the association of the spectra with specific tissue pathologies and comparison of pathology. Most importantly, the exact location of any rupture (the culprit lesion) was not uncovered in the tissue sections. For this reason, the exact nature of the gradients within 10 μm of any tear in the fibrous cap could not be determined. However, the fact that gradients in collagen and elastin similar to those observed in AAAs do exist in the vicinity of a plaque rupture suggests that similar mechanisms of protein degradation may be responsible in both disease states. Thus, an increase in C1 at the expense of C3 (and possibly of elastin) might serve as a marker of plaques needing an immediate intervention.

NIR and IR spectra are distinctive for major proteins in the blood vessel wall (specifically collagens and elastin). Vibrational spectrometry is a potentially useful technique for investigating vascular changes and protein composition associated with cardiovascular disease. Synchrotron IR microspectroscopy has the intensity and spatial resolution needed to quantify collagens and elastin within 10 μm of the site of plaque rupture. These early results support an expanded study in the future employing IR spectra to characterize completely chemical compositions within 10 μm of the location of plaque rupture. If the compositions of many of these small regions prove similar in many different patients, a useful marker will have been demonstrated that can be targeted by NIR catheters *in vivo*.

ACKNOWLEDGMENTS

The authors thank Dr David L. Wetzel for his collaboration on research with atherosclerotic plaque at the National Synchrotron Light Source, Brookhaven National Laboratory.

REFERENCES

1. Glass, C. K. and Witztum, J. L., *Cell* 2001, 104, 503–16.
2. Manning, M. W., *Vasc. Med.* 2002, 7, 45–54.
3. Ross, R., *N. Eng. J. Med.* 1999, 340, 115–26.
4. Ross, R., *Am. Heart J.* 1999, 138, S419–20.
5. Mitchell, M. E. and Sidawy, A. N., *Semin. Vasc. Surg.* 1998, 11, 134–41.
6. Nordestgaard, B. G. and Nielsen, L. B., *Curr. Opin. Lipidol.* 1994, 5, 252–7.
7. Hamilton, C. A., *Pharmacol. Ther.* 1997, 74, 55–72.
8. Williams, K. J. and Tabas, I., *Curr. Opin. Lipidol.* 1998, 9, 471–4.
9. Navab, M., *Arterioscler. Thromb. Vasc. Biol.* 1996, 16, 831–42.
10. Berliner, J., *Thromb. Haemost.* 1997, 78, 195–9.
11. Navab, M., *Coron. Artery Dis.* 1994, 5, 198–204.
12. Schwartz, C. J., Valente, A. J., and Sprague, E. A., *Am. J. Cardiol.* 1993, 71, 9B–14B.
13. Stary, H. C., *Eur. Heart J.* 1990, 11(Suppl E), 3–19.
14. DiCorleto, P. E., *Am. J. Hypertens.* 1993, 6, 314S–18S.
15. Stary, H. C., *Arterioscler. Thromb. Vasc. Biol.* 1995, 15, 1512–31.
16. Campa, J. S., Greenhalgh, R. M., and Powell, J. T., *Atherosclerosis* 1987, 65, 13–21.
17. Powell, J. and Greenhalgh, R. M., *J. Vasc. Surg.* 1989, 9, 297–304.
18. Rizzo, R. J., *J. Vasc. Surg.* 1989, 10, 365–73.

19. Sakalihasan, N., *Eur. J. Vasc. Surg.* 1993, 7, 633–7.
20. Baxter, B. T., *J. Vasc. Surg.* 1994, 19, 797–802; discussion 803.
21. White, J. V. and Mazzacco, S. L., *Ann. N. Y. Acad. Sci.* 1996, 800, 97–120.
22. Ghorpade, A. and Baxter, B. T., *Ann. N. Y. Acad. Sci.* 1996, 800, 138–50.
23. White, J. V., *J. Vasc. Surg.* 1993, 17, 371–80; discussion 380–1.
24. Minion, D. J., *J. Surg. Res.* 1994, 57, 443–6.
25. Halloran, B. G. and Baxter, B. T., *Semin. Vasc. Surg.* 1995, 8, 85–92.
26. Menashi, S., *J. Vasc. Surg.* 1987, 6, 578–82.
27. Dempsey, R. J. and Lodder, R. A., *Appl. Spectrosc.* 1996, 50, 18A–34A.
28. Spielman, A. J., *Brain Res.* 2000, 866, 313–25.
29. Shaw, R. A., *J. Inorg. Biochem.* 2000, 79, 285–93.
30. McKinley, B. A., *J. Trauma* 2000, 48, 637–42.
31. Gabriely, I., *Diabetes Care* 1999, 22, 2026–32.
32. Jaross, W., *Atherosclerosis* 1999, 147, 327–37.
33. Cassis, L. A. and Lodder, R. A., *Anal. Chem.* 1993, 65, 1247–56.
34. Dempsey, R. J., *Ann. N. Y. Acad. Sci.* 1997, 820, 149–69.
35. Moreno, P. R., *Circulation* 2002, 105, 923–7.
36. Daugherty, A., Manning, M. W., and Cassis, L. A., *J. Clin. Invest.* 2000, 105, 1605–12.
37. Bullivant, S., Freeze Etching and Fracturing. In *Some Biological Techniques in Electron Microscopy*, Parsons, D. F. (Ed.), Academic Press: New York, 1970, pp. 123–46.
38. Sharifi, A. M., *J. Hypertens.* 1998, 16, 457–66.
39. Thomas, E. V. and Haaland, A. C., *Anal. Chem.* 1990, 62, 1091–99.
40. Beebe, K. R. and Kowalski, B. R., *Anal. Chem.* 1987, 59, 1007A–17A.
41. Urbas, Aaron, Manning, Michael, W., Daugherty, Alan, Cassis, Lisa, A., Lodder, Robert, A., Near-infrared spectrometry of abdominal aortic aneurysm in the ApoE^{-/-} mouse. *Anal. Chem.* June 3, 2003, 75, 3650–5.
42. Naghavi, M., Libby, P., Falk, E. et al., From vulnerable plaque to vulnerable patient: A call for new definitions. *Circulation* October 2003, 108, 1664–72.
43. Kolodgie, F. D., Burke, A. P., Farb, A. et al., The thin-cap fibroatheroma: A type of vulnerable plaque: The major precursor lesion to acute coronary syndromes. *Curr. Opin. Cardiol.* 2001, 16, 285–92.
44. Wissler, R. W. and Strong, J. P., Risk factors and progression of atherosclerosis in youth. PDAY Research Group. Pathological determinants of atherosclerosis in youth. *Am. J. Pathol.* 1998, 153, 1023–33.
45. Tuzcu, E. M., Kapadia, S. R., Tutar, E. et al., High prevalence of coronary atherosclerosis in asymptomatic teenagers and young adults: Evidence from intravascular ultrasound. *Circulation* 2001, 103, 2705–10.
46. Pasterkamp, G., Schoneveld, A. H., van der Wal, A. C. et al., Relation of arterial geometry to luminal narrowing and histologic markers for plaque vulnerability: The remodeling paradox. *J. Am. Coll. Cardiol.* 1998, 2, 655–62.

35 *In Vivo* and *In Vitro* Near-Infrared Spectroscopic Determination of Blood Glucose and Other Biomedical Components with Chemometrics

*Yi Ping Du, Sumaporn Kasemsumran, Jian-Hui Jiang, and
Yukihiro Ozaki*

CONTENTS

35.1	Introduction	674
35.2	Wavelength Interval Selection in Multicomponent Spectral Analysis	676
35.2.1	Moving Window Partial Least-Squares Regression.....	677
35.2.2	Changeable Size Moving Window Partial Least Squares.....	678
35.2.3	Searching Combination Moving Window Partial Least Squares	679
35.3	<i>In Vivo</i> and <i>In Vitro</i> NIR Spectroscopic Determination of Biomedical Component with MWPLSR, CSMWPLS, and SCMWPLS	681
35.3.1	NIR Spectroscopic Determination of Human Serum Albumin, γ -Globulin, and Glucose in a Control Serum Solution with SCMWPLS	681
35.3.2	Application of SCMWPLS to Noninvasive Blood Glucose Assay with NIR Spectroscopy	686
35.4	Region Orthogonal Signal Correction	689
35.4.1	Orthogonal Signal Correction.....	690
35.4.2	Region Orthogonal Signal Correction	691
35.5	Removal of Interference Signals Due to Water from <i>In Vivo</i> NIR Spectra of Blood Glucose by ROSC	692
35.5.1	Spectra Pretreated by OSC and ROSC	693
35.5.2	Comparison of Prediction Ability between Models Developed by Using OSC- and ROSC-Pretreated Spectra.....	694
35.5.3	Selection of Spectral Regions to Improve PLS Model Performance.....	694
References	697

35.1 INTRODUCTION

In vivo and *in vitro* near-infrared (NIR) spectroscopic determination of blood glucose and other biomedical components with chemometrics have received keen interest because it is a noninvasive analytical technique with no or little pretreatment [1–37]. Particularly, *in vivo* blood glucose measurement has been a matter of great importance since the demand for the noninvasive blood glucose assays has been growing due to the worldwide increase in the number of diabetics [22–26,28,33–37]. The noninvasive blood glucose measurements by NIR spectroscopy may open the possible use for clinical diagnosis, examination for diabetes at a hospital or a house, and have glucose monitoring devices.

NIR spectral measurements of human skin stretch back to 1950s. Real trials for *in vivo* and *in vitro* NIR spectroscopic determination of biomedical components with chemometrics started in 1980s. There have been two major streams for NIR-chemometrics investigations on biomedical components. One is concerned with the quantitative analysis of various biomedical analytes in *in vitro* samples, especially in human serum, and the other is *in vivo* blood glucose measurement for possible application to clinical diagnosis.

As for the first type of explorations, in 1987 Peuchant et al. [4] reported a new approach using NIR reflectance spectroscopy and multiple linear regression (MLR) for measuring cholesterol in human serum samples. From 1992 to 1995, Hall and Pollard [5–7] studied the determination of proteins, triglycerides, and glucose in human sera using NIR spectroscopy with chemometrics, such as least-squares regression, MLR, and partial least-squares (PLS) regression. In 1995, Domjan et al. [8] investigated the determination of protein and β -lipoprotein in human serum using transfection NIR measurements with MLR and PLS methods. Hazen et al. [9] carried out the measurement of glucose and other analytes in 242 undiluted human serum samples by using NIR transmission spectroscopy and PLS regression in 1998. The simultaneous determination of human serum antigen (HSA) and γ -globulin in control serum samples by NIR spectroscopy combined with PLS2 was reported by Murayama et al. in 1998 [10]. A more recent investigation was concerned with the quantitative determination of HSA, γ -globulin, and glucose in control serum samples by using NIR spectroscopy and searching combination moving window partial least squares (SCMWPLS) [11].

For plasma samples, in 1993 Small et al. carried out the NIR spectroscopic measurement of glucose in plasma by coupling digital filtering with PLS regression [12]. In 1998, Heise et al. investigated the multicomponent assay for the blood substrates of total protein, glucose, total cholesterol, triglycerides, and urea in human plasma by using NIR spectroscopy with PLS method [13].

As for NIR studies of whole blood, NIR-PLS determination of glucose in whole blood from four different subjects was studied by Haaland et al. in 1992 [14]. In 1995, Norris and Kuenstner [15] evaluated different multivariate calibration models for several blood components by using transmission NIR measurement. Hazen et al. [16,17] studied the temperature-intensive NIR spectroscopic determination of glucose in aqueous solutions prepared as blood model systems. Glucose in variable matrix of protein and triglyceride diluted by buffer solutions was determined by Pan et al. [18] by using NIR spectroscopy. NIR-PLS determination of HSA and γ -globulin in phosphate buffer solutions were reported by Murayama et al. in 1998 [20]. Recently, Kasemsumran et al. [21] applied a wavelength selection method named moving window partial least-squares regression (MWPLSR) to NIR spectra of HSA, γ -globulin, and glucose in phosphate buffer solutions to build efficient PLS models for their determination.

Several research groups reported their trials for the *in vivo* NIR determination of blood glucose in human skin. In 1992, Robinson et al. [22] reported three different instrument configurations for the measurement of glucose concentration using diffuse transmittance NIR through the finger in the 1600 to 1300 nm region. Heise et al. [25] and Marbach et al. [23,24] carried out intensively the *in vivo* determination of glucose on a single diabetic using glucose tolerance test and that on a population of 133 different subjects. They measured the diffuse reflectance of NIR light at the oral mucosa. In

2003, Maruo et al. [26] reported *in vivo* noninvasive measurement of blood glucose by NIR diffuse reflectance spectroscopy. The NIR spectra of human forearm were measured *in vivo* by using an optimized diffuse reflectance accessory during the oral intake experiments for one diabetic and five healthy subjects [26].

NIR spectroscopy has the following advantages for *in vivo* and *in vitro* determination of blood glucose and other biomedical components with chemometrics. First, it is a nondestructive and noninvasive analytical technique. Second, no or little pretreatment are requested for the NIR analysis. Third, reagents and preprocessing samples are needless. Forth, NIR allows to carry out multicomponent analysis. Moreover, it requires minimal technical expertise.

The noninvasive blood glucose measurement with NIR spectroscopy is a very challenging project because it deals with very weak signals of glucose directly from human skin, and the physiological conditions of skin tissue such as body temperature change easily with time. One critical difficulty associated with *in vivo* blood glucose assay is an extremely low signal-to-noise ratio (S/N) of glucose existing in human tissue. In *in vitro* study using NIR transmittance spectroscopy, if one measures the difference of 10 mg/dl glucose using a 1 mm cell, one will obtain the absorbance change of less than 10 μ AU (absorbance unit) [2]. This means that background noise easily hides the glucose signal when spectra are measured *in vivo*. Therefore, the quality of spectra is critical for reliable assay of blood glucose.

The chemometric method is an essential tool for analyzing complicated NIR spectra of biomedical samples that always show overlapping absorption bands. In the construction of a calibration model, PLS regression is the most popular multivariate method [38,39]. In general, PLS is a powerful method, but when it is applied to NIR spectra of very complex samples consisting of a number of components, it does not always yield good results. This is because such NIR spectra usually contain interface signals, such as those due to water and other components.

The crucial point for building the best calibration models for the determination of blood component concentrations is to select the informative NIR regions where one obtains an optimized calibration model for each component. Hall and Pollard employed the 4926–4200 cm^{-1} region for the determination of total protein, albumin, and globulin by MLR [5], and the combinations of two regions of 7547–5555 and 4914–4210 cm^{-1} for the determination of serum glucose by PLS method [6]. They selected the wavelength regions by considering the correlation plot and correlation spectrum. They compared these plots with the second-derivative of the reflectance spectra of dried samples.

Domjan et al. [8] and Norris and Kuenstner [15] used a similar wavelength selection method for the determination of blood components with NIR transmittance spectroscopy and chemometrics. Domjan et al. [8], selected some wavelengths in the 4950–4202 cm^{-1} region for the quantitative determination of albumin and γ -globulin by MLR method. Norris and Kuenstner [15], chose the region of 6452–5405 cm^{-1} for the determination of albumin by PLS method. Arnold and Small et al. [27], collected the NIR transmission spectra in the 5000–4000 cm^{-1} region of glucose in an aqueous solution and in a biological matrix solution, because glucose in KBr shows their strong absorption bands in this region. They employed a narrow band-pass optical interference filter for extracting a glucose band centered at 4400 cm^{-1} for the quantitative determination of glucose. Heise et al. [13], used some spectral regions for a multicomponent assay of blood components such as total protein, glucose, total cholesterol, triglycerides, and urea in human plasma. These spectral regions were selected based on the NIR absorbance spectra of the crystalline compounds or related substances and their property correlation spectra. They chose the informative region of 6001–5508 cm^{-1} and the combination of the two regions of 6788–5461 and 4736–4212 cm^{-1} to build PLS calibrations for the determination of total protein and serum glucose, respectively. Furthermore, Marbach [19], developed a closed form method that provides a closed-form for statistical calibration models in terms of the pure component spectra, spectra noise and noise in the reference method results, and clues for optimization strategies can be obtained from his application-specific signal/noise ratio. As

an example, attenuated total reflection (ATR) and Fourier transform infrared (FTIR) spectra were used for the *in vitro* measurement of glucose in blood plasma [19].

In this way, the NIR determination of some important blood components by PLS usually uses one of the two spectral regions or their combination. One spectral region is the 7500–5500 cm^{-1} region, which contains bands due to the first-overtones and combinations of the stretching modes of CH, OH, and NH groups. The other is the 5000–4000 cm^{-1} region, which corresponds to the region for the combinations of the fundamentals of CH, OH, and NH groups.

Generally, PLS is an efficient method, but it does not always yield good results when NIR spectra of highly complex samples such as those obtained from *in vivo* glucose measurements are analyzed. Therefore, special chemometric methods are needed in combination with typical PLS calculation to improve model performance. The introduction of advanced chemometric methods is crucial to develop a high-performance calibration model for the noninvasive blood glucose examination with NIR spectroscopy.

Therefore, many NIR studies on quantitative determination of biomedical samples used special chemometric methods [38] such as genetic algorithms (GAs), iterative PLS (IPLS), uninformative variable elimination by PLS (UVE-PLS), and interactive variable selection for PLS (IVS-PLS) in combination with typical PLS calculation to improve model performance. We also proposed two new methods named moving window partial least-squares regression (MWPLSR) [40] and searching combination moving partial least squares (SCMWPLS) [41] as wavelength selection methods and used them together with PLS for the *in vivo* and *in vitro* determination of blood glucose [11,21,30,32,42]. Quite recently, a novel method for pretreatment of NIR spectra called region orthogonal signal correlation (ROSC) was proposed in conjugation with the NIR determination of blood glucose [31]. ROSC was applied to NIR spectra of blood glucose measured *in vivo* for removing an interference signal due to water from the whole spectra [31].

The purpose of this chapter is to review *in vivo* and *in vitro* NIR determination of blood glucose and other biomedical components by chemometrics. This chapter consists of two parts. One part describes newly developed chemometrics methods that are very useful for *in vivo* and *in vitro* NIR spectroscopic determination of blood glucose and the other part is concerned with their applications to the NIR determination of *in vivo*. They involve MWPLSR [40], changeable size moving window partial least squares (CSMWPLS) [41], SCMWPLS [41], and ROSC [31]. MWPLSR, CSMWPLS, and SCMWPLS are concerned with the selection of a wavelength or wavenumber region. MWPLSR is a method to select out informative regions from spectra of a system, CSMWPLS is a method to optimize an informative region, that is, to search for an optimized subregion in a selected informative region, and SCMWPLS searches for the optimized combination of informative regions based on CSMWPLS. Thus, the three methods are linked together.

35.2 WAVELENGTH INTERVAL SELECTION IN MULTICOMPONENT SPECTRAL ANALYSIS

Partial least squares is a full-spectral calibration method and has a built-in capacity to deal with the overdetermined problem of full-spectrum calibration [38,39]. A theoretical demonstration has been given that, under certain assumptions always, the addition of spectral channels always improves the prediction performance [43]. The implication of this proof is that these latent variable (LV) methods may eliminate the necessity of wavelength selection and have a built-in capacity to deal with the overdetermined problem of full-spectrum calibration. However, there is increasing evidence indicating, either theoretically [44,45] or experimentally [46,47], that wavelength selection can still significantly refine the performance of these full-spectrum calibration techniques. Some methods of wavelength or wavenumber selection, such as stepwise selection [48], simplex optimization [49], branch and bound [50], simulated annealing [51], genetic algorithms [52], interval partial least-squares regression, and others [53,54], have been proposed and used. Recently, we have demonstrated that the prediction error of indirect (or inverse) calibration may be inflated by including nonideal

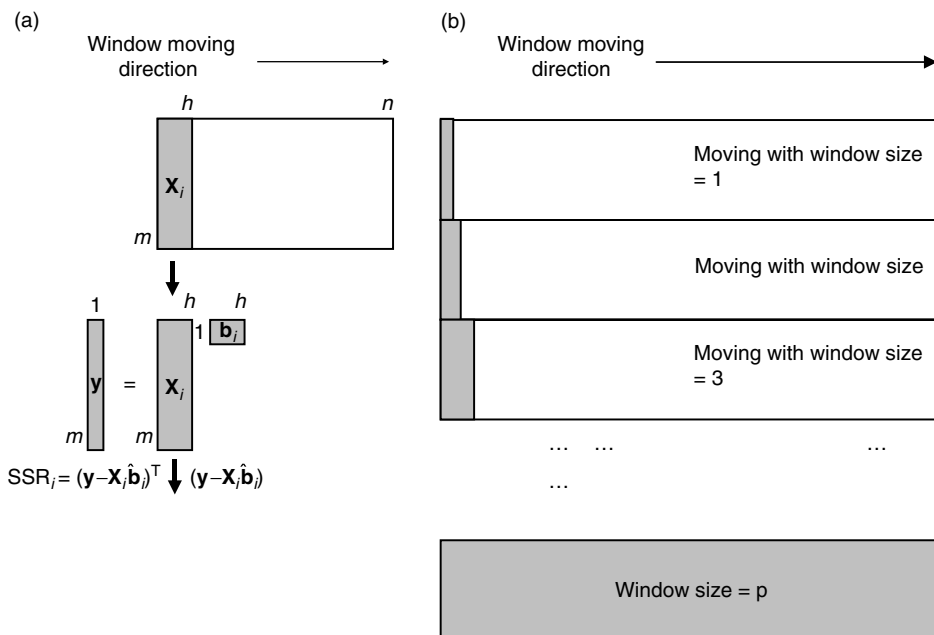


FIGURE 35.1 Scheme for explanation of (a) MWPLSR and (b) CSMWPLS.

spectral regions, and a common feature of the nonideal spectral regions is the increased complexity in LV models when these regions are used for calibration modeling [40]. We, thus, proposed a new wavelength interval selection method, MWPLSR [40]. Then, we further developed CSMWPLSR and SCMWPLS as extension of MWPLSR [41]. MWPLSR, CSMWPLS, and SCMWPLS work together, and they have been actively used for the NIR determination of biomedical components [11,21,30,32,42]. Thus, in this section we focus on MWPLSR, CSMWPLS, and SCMWPLS as to the wavelength selection methods and their applications.

35.2.1 MOVING WINDOW PARTIAL LEAST-SQUARES REGRESSION

The aim of MWPLSR is to search for informative spectral regions for the multicomponent spectral analysis [40]. Informative regions mean that they contain useful information for PLS model building and are helpful to improve the performance of the model. MWPLSR builds a series of PLS models in a window that moves over the whole spectral region and then locates useful spectral intervals, that is, informative regions, in terms of the least complexity of PLS models reaching a desired error level (Figure 35.1a). A salient advantage of MWPLSR is that the calibration model is very stable against the interference from non-composition-related factors. Moreover, the selection of spectral intervals in terms of the least model complexity enables the reduction of the size of a calibration sample set in calibration modeling. An overview of MWPLSR is given below. It is a more detailed description and can be found in Reference [40].

The indirect calibration model routinely used in multicomponent spectral analysis can be described as follows:

$$\mathbf{c} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (35.1)$$

where \mathbf{X} is an $m \times n$ matrix collecting m spectra in rows; each row including n elements, that is, n spectral points; \mathbf{c} , an $m \times 1$ vector of concentration of samples; m and n , the numbers of samples and wavelength or wavenumber points, respectively; \mathbf{e} , the error vector with m elements; and \mathbf{b} , the regression coefficient.

In MWPLSR, a spectral window that starts at the i th spectral channel and ends at the $(I + h - 1)$ th spectral channel is constructed, where h is the window size. The window is moved over the whole

spectral region. There are $(n - h + 1)$ windows over the whole spectra, each window corresponding to a subset of the original spectral \mathbf{X} . One may build PLS models with varying LV numbers from 1 to a fixed value k , and then calculate the sums of squared residuals (SSRs) for each subset. After calculations for all the subsets, SSR is plotted as a function of the position of the window. This will yield a number of residue lines, each line associated with the SSR for a certain LV number in the corresponding window position. A figure containing such residue lines provides two kinds of information. First one is the information about informative regions. Another one is the information about the estimation of LV numbers.

A representative informative region should show low values of SSR and often shows the shape of an upside down peak, which, as will be discussed later, corresponds to a band in the same region. For such peak-shaped region, it is easy to choose the beginning and end points of the region. However, such a selected region possibly does not supply the best predictive results, that is, this region is not optimum, and possibly there still exists a special subregion in this region, which may supply the optimum results. MWPLSR searches for informative regions with a fixed window size. Although it can supply a subregion (a window with the given window size) with the lowest value of SSR in the expected LV number, this subregion is not optimum because the window size is not optimum. Thus, it is necessary to search for the optimum subregion for an informative region obtained by MWPLSR in order to further improve the prediction of the PLS model.

The MWPLSR often chooses more than one informative regions for vibrational spectra data because they contain many spectral bands. In another case when a band of interest is severely interfered with other bands due to some other components in the same system, more than one informative region, which are derived from the separated spectral regions without interference or with not severe interference, may also be selected. In such case, the combination of informative regions seems to be useful to collect more information from the spectra to construct the PLS models. In our previous study, two strategies of combination of regions were suggested, one is the direct combination of all selected regions to build a PLS calibration model, and the other is the combination of the PLS models built individually in each region. The two strategies are not expected to provide optimized results, since the regions are not optimized and the complementarity and correlativity between the regions are not concerned. Therefore, we recently proposed new methods named CSMWPLS and SCMWPLS which enable one to search for an optimized subregion for each selected informative region and the optimized combination of informative regions, respectively. Unlike MWPLSR, CSMWPLS procedure changes the window size and moves the window over a selected informative region with each window size. From all possible spectral intervals within an informative region, CSMWPLS aims at selecting the optimized interval and further improve the performance of the corresponding PLS model. The method of SCMWPLS aims at looking for an optimized combination of informative regions by performing CSMWPLS procedure for every informative region step by step. In this study, MWPLSR is used first to identify informative regions from spectra of a compound, and then we search for an optimized subregion for each selected informative region by CSMWPLS or directly search for the optimized combination of regions by SCMWPLS.

35.2.2 CHANGEABLE SIZE MOVING WINDOW PARTIAL LEAST SQUARES

CSMWPLS is a method to optimize an informative region, that is, to search for an optimized subregion in a selected informative region. The basic idea of CSMWPLS is to change the moving window size w from 1 to p for a given informative region with p spectral points. A moving window is moved from the first spectral point to the $(p - w + 1)$ th point over the informative region and to collect all possible sub-windows for every window size. Figure 35.1b illustrates the algorithm of CSMWPLS. When $w = 1$, moving the window from the first to the end point will collect all possible sub-windows with the window size of one. Similarly, for other cases of w , all sub-windows with the size of w may be obtained. In this way, CSMWPLS considers all possible spectral intervals

(sub-window or subregions) in the range of the given informative region. For every window, a PLS model with a selected LV number is built, and root mean square error of calibration (RMSEC) is calculated. Comparing values of RMSEC for all subregions, the subregion with the smallest value of RMSEC is considered as the optimized spectral interval.

When constructing a PLS model, the selection of the LV number is important. If more variables are selected, the model will easily result in overfitting, while the selection of fewer variables will cause underfitting. Cross validation is a prediction-based variable selection method, in which some samples are picked out for calculating predictive residual error sum of squares (PRESS) [55]. PRESS is a commonly used criterion for LV number selection. Leave-one-out cross validation method is of most application to choose the number of variables in chemometrics. However, Martens and Dardenne [56] pointed out that leave-one-out cross validation often causes overfitting because only one sample is left out. Accordingly, if the number of samples in the data set used is large enough, more samples are suggested to leave out [57,58].

A rational subregion selected by CSMWPLS should develop such a PLS model that the RMSEC of the model is expected to reach an acceptable error level with a reasonable LV number. In CSMWPLS for simplicity and automatic selection by a computer, the subregion only with the smallest value of RMSEC is considered as the optimum and selected. In this situation, CSMWPLS selects quite easily a subregion with not only the smallest RMSEC, but also an unreasonable LV number. Therefore, in CSMWPLS, a maximum LV number is constrained to avoid selecting unreasonable LV number, say, the selected LV number by cross validation must not be larger than the maximum. LV number of the PLS model for an informative region can be easily estimated by regressing the spectra in the region against the concentration c . The LV number is determined to be the number where RMSEC starts to decrease insignificantly with the increase of LV number. This LV number is considered as the maximum LV number.

35.2.3 SEARCHING COMBINATION MOVING WINDOW PARTIAL LEAST SQUARES

SCMWPLS is a novel method to search for the optimized combination of informative regions based on CSMWPLS. Every informative region selected by MWPLSR, as discussed above, contains an optimized subregion, which can be found by CSMWPLS. Although these subregions are optimum in their corresponding regions, their direct combination possibly cannot show an optimum performance. Therefore, SCMWPLS was developed as a novel method to search for the optimized combination of informative regions.

Exhaustive search, which considers all possible combinations of special points in all informative regions, guarantees to find out the global optimum. However, there will be huge number of combinations even if the numbers of spectral points in the informative regions are relatively small. Therefore, the calculation by exhaustive search is not recommended.

Since rational vibrational spectra show continuity of spectral bands, we only collect the combinations of continuous spectral points, that is, spectral intervals, but not discrete spectral points in every region. This will result in a significant decrease in the number of combinations. However, in general, it is still too large to try the exhaustive search. Let us consider a system with three informative regions A, B, and C, where the numbers of spectral intervals are a , b , and c , respectively. For this system the number of possible combinations equals to

$$\frac{a(a+1)}{2} \times \frac{b(b+1)}{2} \times \frac{c(c+1)}{2}$$

If $a = 40$, $b = 50$, and $c = 60$, one has

$$\frac{40(40+1)}{2} \times \frac{50(50+1)}{2} \times \frac{60(60+1)}{2} = 1.9133 \times 10^9$$

combinations in exhaustive search method.

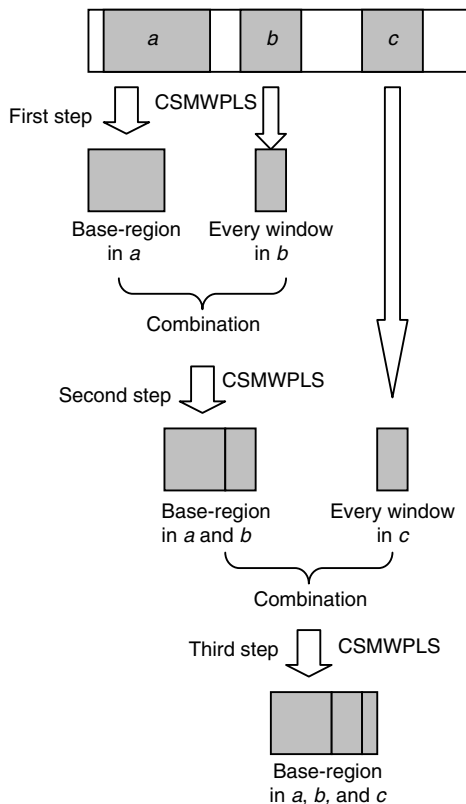


FIGURE 35.2 Scheme for explanation of SCMWPLS. (Reproduced from Y.P. Du, Y.Z. Liang, J.H. Jiang, R.J. Berry and Y. Ozaki, *Anal. Chim. Acta*, **501**, 183 (2004). With permission. Copyright (2004) Elsevier.)

Du et al. proposed a local optimized algorithm to search for the optimized combination of informative regions. Figure 35.2 shows a scheme for explanation of SCMWPLS. The first step of SCMWPLS is, to apply CSMWPLS to the first region (region a in Figure 35.2) that shows the minimum residue level, to research for the optimized subregion with the smallest value of RMSEC in a reasonable LV number selected by cross validation. This optimized subregion is viewed as the base-region. The second step is to apply CSMWPLS to the second informative region (region b in Figure 35.2), where one uses the combinations of the base-region and one of the possible spectral intervals selected from the second informative region, to build PLS models and calculate their RMSEC values. By the second step, a new base-region with the smallest value of RMSEC will be chosen. The next step is to look for another new base-region with the similar procedure for the next informative region, until the last informative region. The final base-region after finishing calculations for all the informative regions is considered as the optimized combination. This algorithm, unlike exhaustive search, only needs to search for a few parts of all possible combinations of the informative regions, whose number is

$$\frac{a(a+1)}{2} + \frac{b(b+1)}{2} + \frac{c(c+1)}{2}$$

for the system with three informative regions A, B, and C, where the numbers of spectral intervals are a , b , and c , respectively. When $a = 40$, $b = 50$, and $c = 60$, SCMWPLS will deal with

$$\frac{40(40+1)}{2} + \frac{50(50+1)}{2} + \frac{60(60+1)}{2} = 3925$$

combinations. Note that the number of combinations decreases remarkably, compared with that of exhaustive search (1.9133×10^9).

One must consider two points when SCMWPLS is employed. The first one is the selection of LV number of a PLS model. One can use the same strategy of the selection as that for CSMWPLS. Since, in general, there is more than one informative region, the most informative region, which should show almost the smallest value of RMSEC and expected LV number, will be ascertained first. The maximum LV number, constrained to the algorithm (as in the case of CSMWPLS) is chosen according to the LV number of the PLS model including the most informative region. The second point is the order of informative regions to search for. Usually, the value of RMSEC for an expected LV number of a PLS model including an informative region reflects the importance of the region. Therefore, the region with the smallest RMSEC, that is, the most informative region, should be the first in the order, and then the region with the second smallest RMSEC, and so on. Even so, if possible, all the possible orders are suggested to attempt, for example, for the three informative regions A, B, and C, it is best to try all the orders, A–B–C, A–C–B, B–A–C, B–C–A, C–A–B, and C–B–A.

35.3 IN VIVO AND IN VITRO NIR SPECTROSCOPIC DETERMINATION OF BIOMEDICAL COMPONENT WITH MWPLSR, CSMWPLS, AND SCMWPLS

35.3.1 NIR SPECTROSCOPIC DETERMINATION OF HUMAN SERUM ALBUMIN, γ -GLOBULIN, AND GLUCOSE IN A CONTROL SERUM SOLUTION WITH SCMWPLS

Near-infrared spectra in the region of 5000 to 4000 cm^{-1} with SCMWPLS were employed to determine the concentrations of human serum albumin (HSA), γ -globulin, and glucose contained in the control serum IIB (CS IIB) solutions with various concentrations [11]. PLS calibration models using the regions obtained by SCMWPLS were developed for HSA, γ -globulin, and glucose. These models yielded good prediction with the smallest RMSEP, the relatively small number of PLS factors, and the highest correlation coefficients among the results achieved by using the whole region and the informative region selective by MWPLSR [11]. The RMSEP values of HSA, γ -globulin, and glucose yielded by SCMWPLS were 0.0303, 0.0327, and 0.0195 g/dl, respectively. These results prove that SCMWPLS can be successfully applied to determine simultaneously the concentrations of HSA, γ -globulin, and glucose in complicated biological fluids such as CS IIB solutions by using NIR spectroscopy [11].

The CS IIB solution contained various kinds of biological metabolites, such as total cholesterol (241 mg/dl), triglyceride (172 mg/dl), bilirubin (6.50 mg/dl), urea N (51.6 mg/dl), total protein (5.40 g/dl), and so forth. Nevertheless, the concentration of each component was always kept constant. 3.20 g/dl HSA, 2.20 g/dl γ -globulin, and 0.28 g/dl glucose were contained in the CS IIB solution. Stock solutions of HSA, γ -globulin, and glucose were prepared by dissolving HSA, γ -globulin, and glucose powder into the CS IIB solutions, respectively. The 125 mixture samples were prepared by mixing appropriate volumes of the stock solutions and CS IIB solution with fixing the concentrations of two stock solutions and varying the concentration of another stock solution. All these concentrations were designed to be five levels of full factorial. Five concentration levels of HSA, γ -globulin, and glucose were (3.20, 3.90, 4.60, 5.30, and 6.00 g/dl), (2.20, 2.65, 3.10, 3.55, and 4.00 g/dl), and (0.28, 0.71, 1.14, 1.57, and 2.00 g/dl), respectively.

The original spectra were subjected to multiplicative scatter correction (MSC) before applying MWPLSR and SCMWPLS and multivariate analysis. The spectral data of 125 samples were divided randomly into three sets: calibration, validation, and prediction sets contain 80, 15, and 30 samples, respectively.

Figure 35.3 shows a NIR spectrum in the 12,000–4,000 cm^{-1} region of control serum sample. A major component of control serum is water [11]. An intense band near 6900 cm^{-1} is assigned to the combination of OH symmetric and antisymmetric stretching modes of water, and a saturated feature around 5150 cm^{-1} is attributed to the combination of its OH stretching and bending vibrations. To avoid interference from the strong absorption bands of water, the 5000–4000 cm^{-1} region was used for searching the informative regions of HSA, γ -globulin, and glucose. Figure 35.4 displays normalized NIR spectra in the region of 8000–4000 cm^{-1} of water and powder reagents of glucose, γ -globulin, and HSA [11]. These spectra reveal that the three biomedical samples show high absorbance in the 5000–4000 cm^{-1} region while water yields relatively weak absorbance in the same region. In this region many bands due to the combinations of stretching and deformation modes of CH, OH, and NH groups appear. Therefore, this region is very suitable for collecting useful information about the three analytes and developing PLS models.

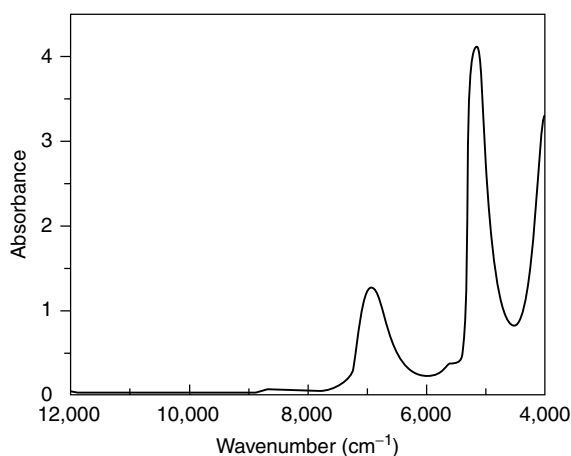


FIGURE 35.3 A NIR spectrum of CS IIB solution.

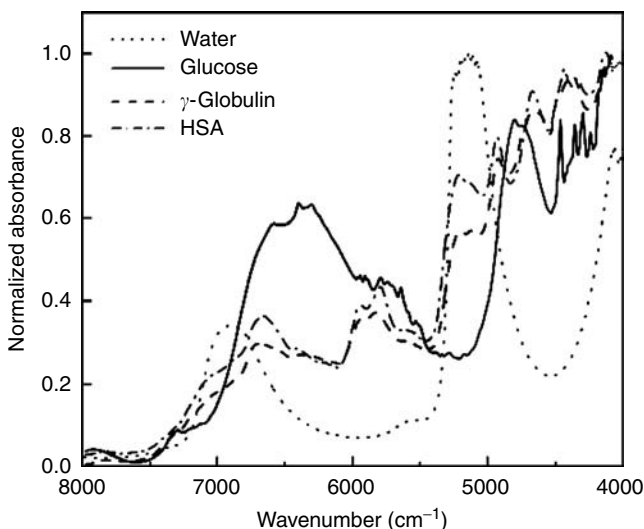


FIGURE 35.4 NIR normalized absorbance spectra in the 8000 to 4000 cm^{-1} region for water, glucose powder, γ -globulin powder, and HSA powder. (Reproduced from S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne and Y. Ozaki, *Anal. Chim. Acta*, **512**, 223 (2004). With permission. Copyright (2004) Elsevier.)

Figure 35.5 to Figure 35.7 show the residue lines for HSA, γ -globulin, and glucose obtained by MWPLSR for the NIR data in the spectral region of 5000–4000 cm^{-1} , respectively [11]. It is noted in Figure 35.5 that the spectral regions of 4800–4500 and 4490–4260 cm^{-1} show relatively small values of SSR (residue lines) and are considered as informative regions. Figure 35.6 provides the expected informative regions of 4800–4617, 4606–4502, 4489–4412, and 4399–4250 cm^{-1} of γ -globulin. These regions correspond to the regions for the combinations of the stretching and deformation modes of CH_2 , CH_3 , NH_2 , and amide groups of the protein. Figure 35.7 shows four

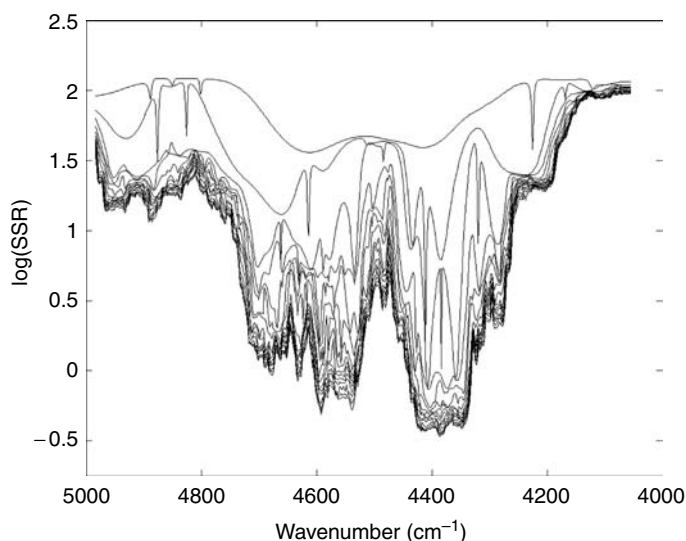


FIGURE 35.5 Residue lines obtained by MWPLSR for the NIR spectra of HSA in the region of 5000 to 4000 cm^{-1} . (Reproduced from S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne and Y. Ozaki, *Anal. Chim. Acta*, **512**, 223 (2004). With permission. Copyright (2004) Elsevier.)

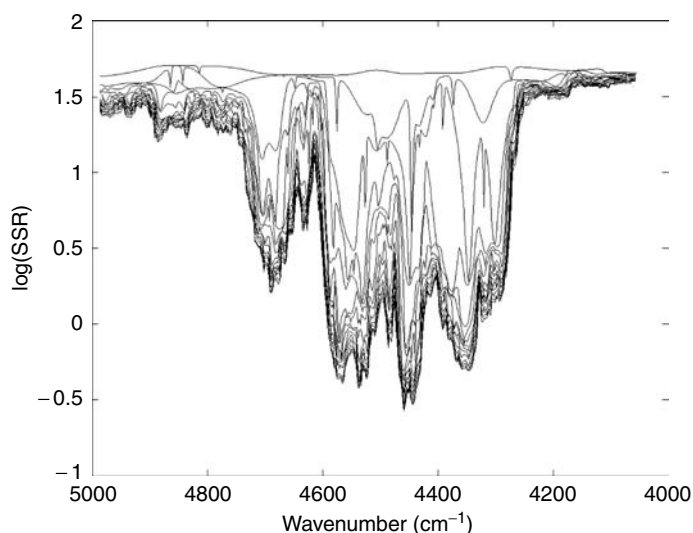


FIGURE 35.6 Residue lines obtained by MWPLSR for the NIR spectra of γ -globulin in the region of 5000 to 4000 cm^{-1} . (Reproduced from S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne and Y. Ozaki, *Anal. Chim. Acta*, **512**, 223 (2004). With permission. Copyright (2004) Elsevier.)

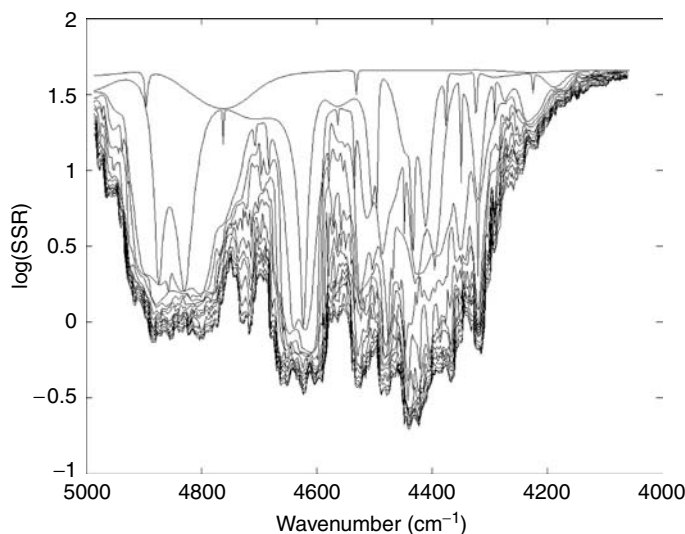


FIGURE 35.7 Residue lines obtained by MWPLSR for the NIR spectra of glucose in the region of 5000 to 4000 cm^{-1} . (Reproduced from S. Kasemsunran, Y. P. Du, K. Murayama, M. Huehne and Y. Ozaki, *Anal. Chim. Acta*, **512**, 223 (2004). With permission. Copyright (2004) Elsevier.)

expected informative regions of glucose. The regions of 4930–4756, 4744–4704, 4685–4571, and 4546–4220 cm^{-1} involve the combinations of CH stretching and deformation modes of glucose [11].

SCMWPLS algorithm was applied to the informative regions of HSA, γ -globulin, and glucose obtained by MWPLSR to search for the optimized combinations of those regions. The prediction results for such obtained combinations were calculated and are listed for the three components in Table 35.1 to Table 35.3, respectively [11]. For comparison, the prediction results of PLS models based on the whole region, each individual informative region and the direct combinations are also listed in the same tables.

For HSA, two individual informative regions are selected by MWPLSR (Table 35.1) [11]. The region of 4800–4500 cm^{-1} gives the better prediction results with the RMSEP of 0.0384 g/dl at PLS factor 7 than the 4490–4260 cm^{-1} region. The direct combination of informative regions cannot provide better prediction results than the individual region of 4800–4500 cm^{-1} , probably because the direct combination region was not optimized. SCMWPLS for the two informative regions of 4800–4500 and 4490–4260 reveals that the combination of 4797–4500 and 4403–4293 cm^{-1} regions provides the best prediction results with the lowest RMSEP of 0.0303 g/dl and PLS factor 6.

For γ -globulin, the prediction results for the PLS model developed by using the whole region are relatively good with the RMSEP of 0.0655 g/dl and PLS factor 10 (Table 35.2). Four informative regions are selected by MWPLSR (see Figure 35.6). The prediction results of PLS models achieved by using the four informative regions and their two direct combinations that show the best results among all possible combinations of the four informative regions are given in Table 35.2. One can find from Table 35.2 that the direct combination of the 4606–4502, 4489–4412, and 4399–4250 cm^{-1} regions yields a good prediction model with the RMSEP of 0.0428 g/dl and PLS factor of 6 and note that SCMWPLS selects the optimized combination of the 4789–4619, 4594–4502, 4478–4472, and 4357–4287 cm^{-1} regions that provides a very good prediction result with the RMSEP of 0.0327 g/dl, PLS factor of 6, and correlation coefficient of 0.9985. These results are significantly better than those calculated by using the individual informative regions and their direct combinations.

It can be seen from Table 35.3 the direct combinations of informative regions such as the combination of 4744–4704, 4685–4571, and 4546–4220 cm^{-1} regions show significant improvements compared with the whole region and individual informative regions. SCMWPLS provides the

TABLE 35.1

Prediction Results for PLS Calibration Models of HSA in CS II Solutions for the Spectral Regions Selected by MWPLSR and SCMWPLS

Method	Spectral region (cm ⁻¹)	PLS factor	Correlation coefficient	RMSEP (g/dl)
Whole region	5000–4017	10	.9950	0.1058
MWPLSR	4800–4500	7	.9993	0.0384
MWPLSR	4490–4260	6	.9987	0.0569
MWPLSR	4800–4500, 4490–4260	7	.9988	0.0475
SCMWPLS	4797–4500, 4403–4293	6	.9996	0.0303

Source: S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne and Y. Ozaki, *Anal. Chim. Acta*, **512**, 223 (2004).

TABLE 35.2

Prediction Results for PLS Calibration Models of γ -Globulin in CS II Solutions for the Spectral Regions Selected by MWPLSR and SCMWPLS

Method	Spectral region (cm ⁻¹)	PLS factor	Correlation coefficient	RMSEP (g/dl)
Whole region	5000–4017	10	.9941	0.0655
MWPLSR	4800–4617	13	.9864	0.0977
MWPLSR	4606–4502	8	.9938	0.0680
MWPLSR	4489–4412	10	.9931	0.0708
MWPLSR	4399–4250	7	.9916	0.0778
MWPLSR	4800–4617, 4606–4502, 4489–4412, 4399–4250	6	.9968	0.0483
MWPLSR	4606–4502, 4489–4412, 4399–4250	6	.9975	0.0428
SCMWPLS	4789–4619, 4594–4502, 4478–4472, 4357–4287	6	.9985	0.0327

Source: S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne and Y. Ozaki, *Anal. Chim. Acta*, **512**, 223 (2004).

TABLE 35.3

Prediction Results for PLS Calibration Models of Glucose in CS II Solutions for the Spectral Regions Selected by MWPLSR and SCMWPLS

Method	Spectral region (cm ⁻¹)	PLS factor	Correlation coefficient	RMSEP (g/dl)
Whole region	5000–4017	9	.9966	0.0549
MWPLSR	4930–4756	5	.9911	0.0970
MWPLSR	4744–4704	8	.9568	0.1917
MWPLSR	4685–4571	5	.9937	0.0894
MWPLSR	4546–4220	4	.9927	0.0870
MWPLSR	4930–4756, 4744–4704, 4685–4571, 4546–4220	4	.9987	0.0345
MWPLSR	4930–4756, 4685–4571, 4546–4220	4	.9991	0.0302
MWPLSR	4744–4704, 4685–4571, 4546–4220	7	.9992	0.0251
SCMWPLS	4895–4764, 4719–4717, 4650–4602, 4546–4256	5	.9995	0.0195

Source: S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne and Y. Ozaki, *Anal. Chim. Acta*, **512**, 223 (2004).

optimized combination of 4895–4764, 4719–4717, 4650–4602, and 4546–4256 cm^{-1} regions. This optimized combination further improves the prediction results, yielding the RMSEP of 0.0195 g/dl, PLS factor of 5, and correlation coefficient of 0.9995.

The results in Table 35.1 to Table 35.3 for the PLS calibrations of HSA, γ -globulin, and glucose in the CS IIB solutions demonstrate the potential of SCMWPLS [11]. These SCMWPLS results always yield the best prediction results for the calibrations of HSA, γ -globulin, and glucose in the CS IIB solutions (Table 35.1 to Table 35.3).

35.3.2 APPLICATION OF SCMWPLS TO NONINVASIVE BLOOD GLUCOSE ASSAY WITH NIR SPECTROSCOPY

We applied SCMWPLS to the blood glucose assay by *in vivo* NIR spectra of human skin [42]. The NIR spectra in the region of 1212–1889 nm were measured for an oral glucose tolerance test. With the use of the optimized informative region selected by SCMWPLS, PLS models were built for the quantitative determination of blood glucose [42]. Root mean squares error of validation (RESEV) and PLS factor numbers were calculated, and error grid analysis (EGA) was given for the performance testing of PLS models developed. The results showed that SCMWPLS can be successfully used in the noninvasive NIR measurements for providing more powerful PLS models than those obtained by the conventional PLS and MWPLSR [42].

The subject was a healthy man. The *in vivo* measurements of NIR spectra of his skin at forearm were started 20 min before the drinking of a 225 ml glucose solution containing 50 g of glucose (Toleran G, Shimizu Pharmaceutical, Japan) to change his blood glucose concentration [26,33]. The NIR spectra were measured noninvasively every 5 min during the oral glucose intake experiment. Figure 35.8 shows a profile of blood glucose concentration measured during the oral glucose tolerance cycle.

The original spectra were subjected to MSC before MWPLSR, SCMWPLS, and multivariate analysis were applied. All 48 skin spectra were employed to build PLS calibration models. The model performance was validated by use of the four segments cross-validation method (12 spectra per segment) and the RMSEV was calculated.

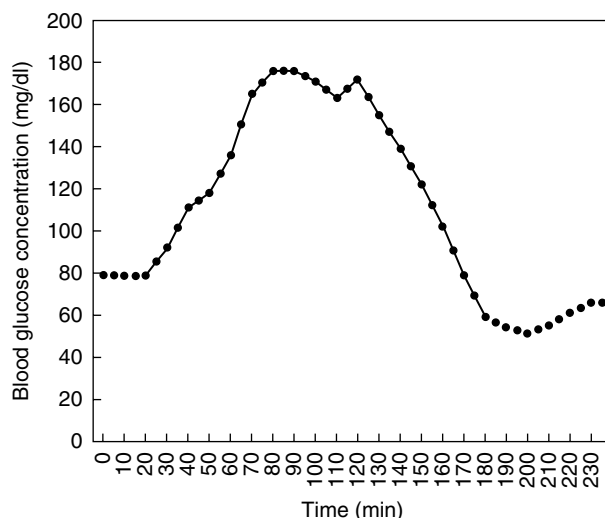


FIGURE 35.8 A profile of blood glucose concentration measured during an oral glucose tolerance test. (Reproduced from S. Kasemsumran, Y. P. Du, K. Maruo and Y. Ozaki, *Chemom. Intell. Lab. Syst.*, **82**, 97 (2006). With permission. Copyright (2005) Elsevier.)

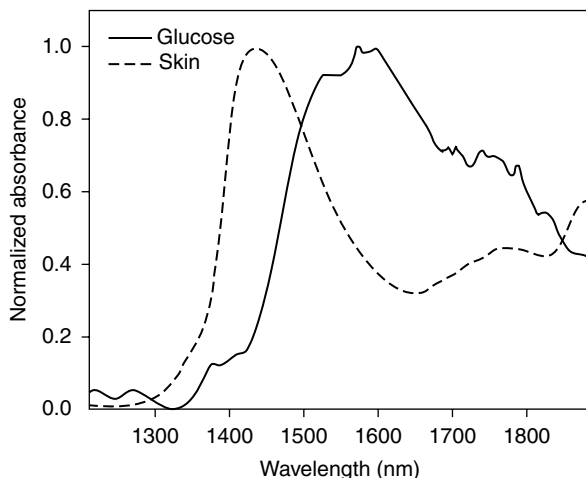


FIGURE 35.9 A normalized NIR spectra in the 1212–1889 nm region of glucose powder and a normalized averaged of 48 human skin spectra measured during an oral glucose tolerance test. (Reproduced from S. Kasemsumran, Y. P. Du, K. Maruo and Y. Ozaki, *Chemom. Intell. Lab. Syst.*, **82**, 97 (2006). With permission. Copyright (2005) Elsevier.)

Figure 35.9 shows normalized NIR spectra of glucose powder (solid line) and skin (dash line) [42]. The latter spectrum was calculated by averaging of 48 skin spectra. The skin spectrum interfered severely by strong absorbance of water. Bands arising from the first overtones of CH stretching modes of glucose are expected to appear in the 1530–1860 nm region. It is noted that absorption features of glucose and other components contained in the skin such as fat tissues cannot be identified directly in the *in vivo* skin spectra. However, the features due to other components may interfere significantly with those due to blood glucose. Wavelength selection methods are very useful to deal with such problems of the interferences. Therefore, MWPLSR and SCMWPLS were applied to the *in vivo* NIR spectra for the blood glucose determination.

Figure 35.10 displays 15 residue lines for blood obtained by applying MWPLSR to the NIR spectra of skin [42]. Three informative regions, the 1228–1323, 1574–1736, and 1739–1800 nm regions, can easily be found. The informative regions of 1574–1736 and 1739–1800 nm contain bands due to the first overtones of OH and CH stretching modes of glucose. The region of 1228–1323 nm contains only weak absorption feature due to glucose but it has weak interference from water.

SCMWPLS algorithm was applied to these informative regions for optimizing the combination of informative regions. It is noted that the optimized combination obtained by SCMWPLS contains only one informative region, that is, the 1574–1736 nm region.

Statistical results of blood glucose models built by use of the whole region, the individual informative regions, and their direct combinations, and the optimized informative region are compared in Table 35.4. The PLS calibration model obtained by using the whole region of 1212–1889 nm yields the large RMSEV of 20.1977 mg/dl with a high PLS factor of 7 and the correlation coefficient of 0.8936. The informative region of 1574–1736 nm gives the best validation results among the three informative regions with the correlation coefficient of 0.9091 and the RMSEV of 18.3642 mg/dl with the PLS factor 4. The direct combinations of informative regions cannot supply better validation results than the individual informative region of 1574–1736 nm. It can be seen from Table 35.4 that the PLS calibration model based on SCMWPLS using the best optimized informative region of 1616–1733 nm yields the best validation results with the highest correlation coefficient of 0.9205 and the lowest RMSEV of 17.1924 mg/dl with the PLS factor 4 [42]. The significant improvement of the validation results shows that this optimized region selected by SCMWPLS contains larger information about blood glucose and less interference than others [42].

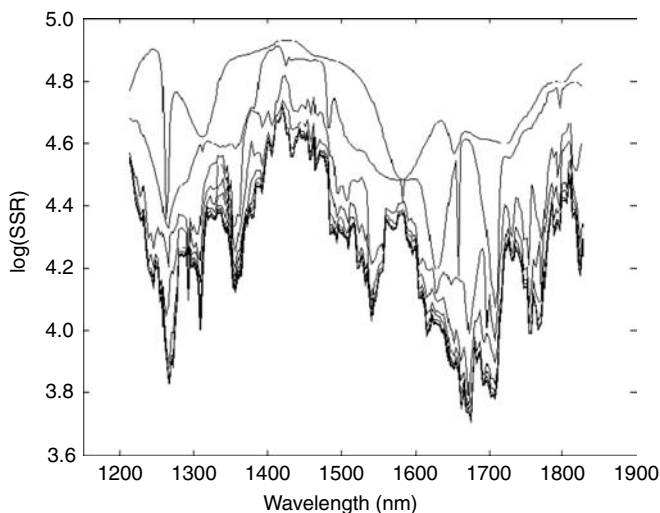


FIGURE 35.10 Residue lines obtained by MWPLSR for the NIR spectra of skin. (Reproduced from S. Kasemsumran, Y. P. Du, K. Maruo and Y. Ozaki, *Chemom. Intell. Lab. Syst.*, **82**, 97 (2006). With permission. Copyright (2005) Elsevier.)

TABLE 35.4

Prediction Results for PLS Calibration Models for Blood Glucose Determination Developed by Use of the Whole Spectral Region and the Regions Selected by MWPLSR and SCMWPLS

Method	Spectral region (nm)	PLS factor	Correlation coefficient	RMSEV (mg/dl)
Whole region	1212–1889	7	.8936	20.1977
MWPLSR	1228–1323	6	.8519	24.2398
MWPLSR	1574–1736	4	.9091	18.3642
MWPLSR	1739–1800	4	.8302	24.8947
MWPLSR	1228–1323, 1574–1736, 1739–1800	6	.8840	20.9073
MWPLSR	1228–1323, 1574–1736	6	.8984	19.4118
MWPLSR	1574–1736, 1739–1800	5	.9060	18.7775
SCMWPLS	1616–1733	4	.9205	17.1924

Source: S. Kasemsumran, Y. P. Du, K. Maruo and Y. Ozaki, *Chemom. Intell. Lab. Syst.*, **82**, 97 (2006).

To confirm the high performance of the PLS model obtained by SCMWPLS, EGA was applied to the three models developed by using the whole region, the region of 1574–1736 nm suggested by MWPLSR, and the optimized region of 1616–1733 nm yielded by SCMWPLS. EGA is useful to judge the performance of the calibration models from the point of clinical accuracy. Figure 35.11a–c show EGA plots for the three PLS models [42]. The plots were obtained by plotting the predicted blood glucose concentrations vs. the actual blood glucose concentrations. All these EGA plots show that the mainstream of the prediction values is located within Zone A that is defined as the clinically correction. These EGA plots confirm that the prediction results for the noninvasive blood glucose measurements by NIR spectroscopy are clinically acceptable. However, the EGA plots for those three models are significantly different from each other in terms of the numbers and positions of points in each zone. Figure 35.12 compares the differences among the three EGA plots. The bar

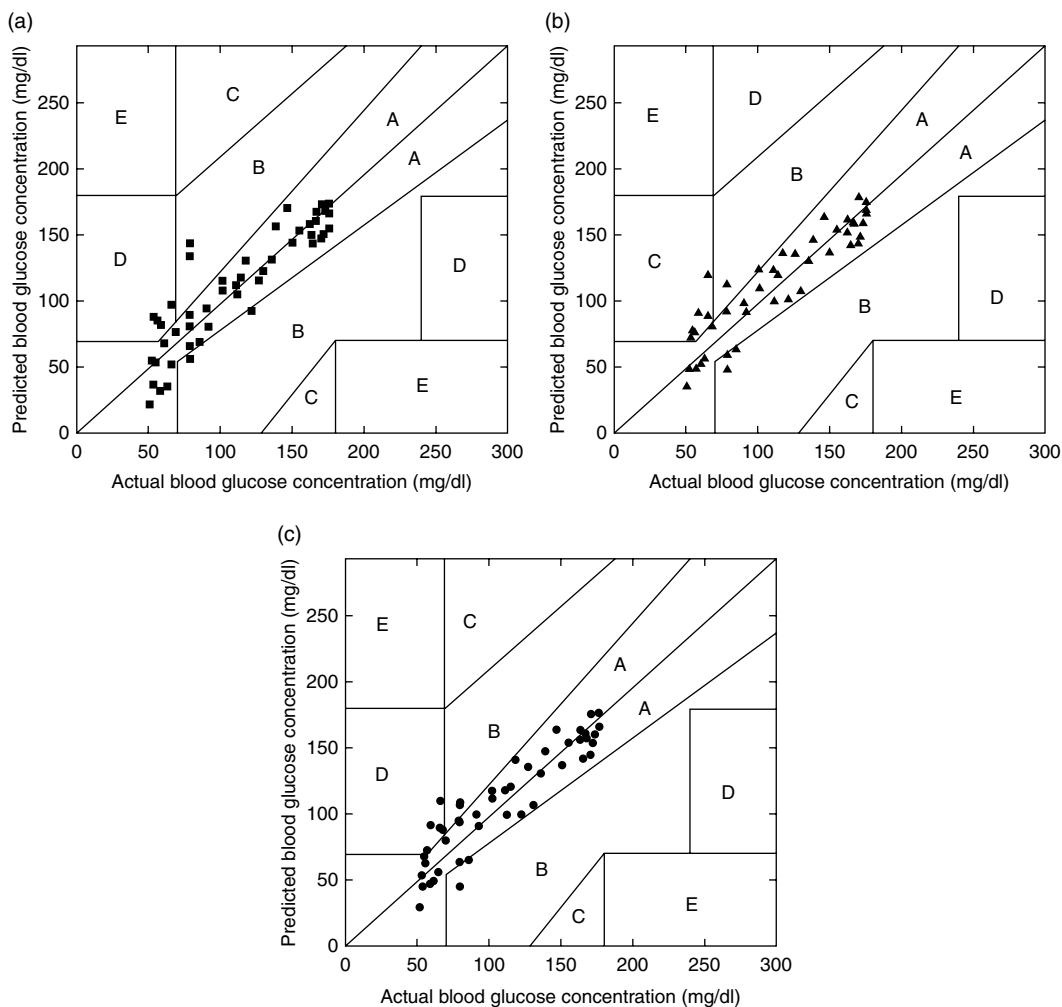


FIGURE 35.11 EGA plots between the actual and predicted blood glucose concentrations for (a) the PLS model based on the whole region, (b) the PLS model obtained by using the informative region of 1574 to 1736 nm suggested by MWPLSR, and (c) the PLS model developed by use of the optimized informative region of 1616 to 1733 nm obtained by SCMWPLS. (Reproduced from S. Kasemsumran, Y. P. Du, K. Maruo and Y. Ozaki, *Chemom. Intell. Lab. Syst.*, **82**, 97 (2006) with the permission. Copyright (2005) Elsevier.)

plot for the EGA results in Figure 35.12 shows that 85.5, 80.2, and 83.3% of the predicted blood glucose concentrations fall in Zone A for the models obtained with SCMWPLS (dark gray color bar), MWPLSR (gray color bar), and the whole region (blank bar), respectively [42]. 7.2, 9.4, and 8.3% of the predicted values are found in Zone B, and 7.3, 10.4, and 8.4% are seen in Zone D for these three models. The EGA results clearly reveal that the PLS model obtained by using the optimized region selected by SCMWPLS provides not only the best performance in statistical point, but also present the best clinical accuracy among the three models.

35.4 REGION ORTHOGONAL SIGNAL CORRECTION

Wold et al. [59] introduced a chemometric method named orthogonal signal correction (OSC) in 1998. The purpose of OSC is to remove strong structured variation (OSC components) from spectra

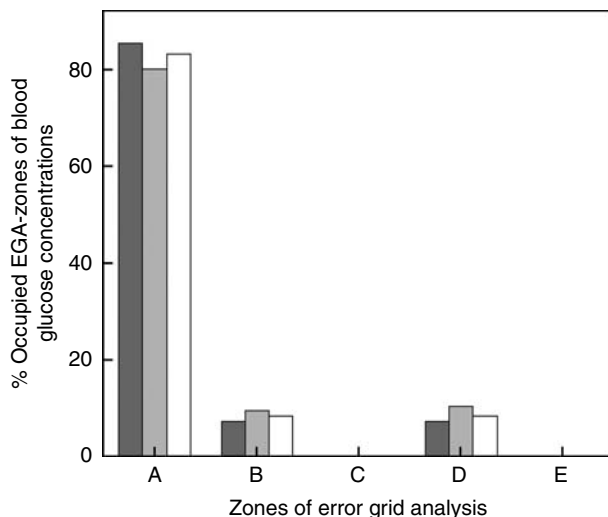


FIGURE 35.12 The plot of predicted blood glucose values found in Zone A, B, C, D, and E for the PLS model built by use of the optimized informative region of 1616 to 1733 nm obtained by SCMWPLS (dark gray color bar), the PLS model based on the informative region of 1574 to 1736 nm suggested by MWPLSR (gray color bar), and the PLS model developed by employing the whole region (blank bar). (Reproduced from S. Kasemsumran, Y. P. Du, K. Maruo and Y. Ozaki, *Chemom. Intell. Lab. Syst.*, **82**, 97 (2006). With permission. Copyright (2005) Elsevier.)

that is not correlated to concentration, that is, is orthogonal to concentration. Since the introduction of OSC, a number of different OSC algorithms have been developed to improve the Wold's OSC procedure [60–62]. Although these algorithms have different calculation procedures and slightly different calculation results, their basic idea is the same. We recently proposed ROSC [31]. ROSC uses a special region of spectra to estimate the variations in interference components and remove these components for other regions, while OSC only uses one fixed region of spectra to calculate OSC components and remove them in the same region. In this region, not only the interference components, but also the interested components are contained. A clear advantage of ROSC is that the orthogonal components estimated by ROSC is more interpretable than OSC components obtained by OSC because one can select a spectral region to remove signals of a special component such as water [31].

35.4.1 ORTHOGONAL SIGNAL CORRECTION

It is assumed that a $m \times n$ matrix \mathbf{X} collecting spectra in rows of m samples and a $m \times 1$ vector \mathbf{y} collecting concentrations of m samples are column mean centered. A subspace can be calculated by $\mathbf{w} = \mathbf{X}^T \mathbf{y}$. Here, the superscript^T denotes transpose. A projection operator \mathbf{w}_{orth} associated with the orthogonal complement of \mathbf{w} may be expressed as $\mathbf{W}_{\text{orth}} = \mathbf{I} - \mathbf{w}(\mathbf{w}^T \mathbf{w})^{-1} \mathbf{w}^T$. Here, \mathbf{I} is an identity matrix. It is proved that any column vector in the matrix \mathbf{X}_{orth} obtained by projecting \mathbf{X} to \mathbf{w}_{orth} is orthogonal to \mathbf{y} : $\mathbf{X}_{\text{orth}} = \mathbf{X} \mathbf{W}_{\text{orth}} = \mathbf{X} - \mathbf{X} \mathbf{w}(\mathbf{w}^T \mathbf{w})^{-1} \mathbf{w}^T$; $\mathbf{y}^T \mathbf{X}_{\text{orth}} = \mathbf{y}^T \mathbf{X} - \mathbf{y}^T \mathbf{X} \mathbf{w}(\mathbf{w}^T \mathbf{w})^{-1} \mathbf{w}^T = \mathbf{w}^T - (\mathbf{w}^T \mathbf{w})(\mathbf{w}^T \mathbf{w})^{-1} \mathbf{w}^T = \mathbf{w}^T - \mathbf{w}^T = \mathbf{0}$ (note, $\mathbf{y}^T \mathbf{X} = (\mathbf{X}^T \mathbf{y})^T = \mathbf{w}^T$). Obviously, any column vector in \mathbf{X}_{orth} is \mathbf{y} -orthogonal and is a possible score vector \mathbf{t}_{orth} to construct OSC component. In OSC, a score vector \mathbf{t}_{osc} that accounts for the largest variation in \mathbf{X}_{orth} is selected. Therefore, singular value decomposition is performed for the matrix \mathbf{X}_{orth} , $\mathbf{X}_{\text{orth}} = \mathbf{U} \mathbf{S} \mathbf{V}^T$, and the first p columns in \mathbf{U} are considered as the scores \mathbf{T}_{osc} of p OSC components, $\mathbf{T}_{\text{osc}} = [\mathbf{t}_{\text{osc},1}, \mathbf{t}_{\text{osc},2}, \dots, \mathbf{t}_{\text{osc},p}] = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p]$. Clearly, \mathbf{T}_{osc} is orthogonal to \mathbf{y} , because of the orthogonality between \mathbf{X}_{orth} and \mathbf{y} , and \mathbf{T}_{osc} shows the feature of being column orthonormality, that is, $\mathbf{T}_{\text{osc}}^T \mathbf{T}_{\text{osc}} = \mathbf{I}$. Thus, loadings of

p OSC components will easily be calculated by least square method, $\mathbf{P}_{\text{osc}} = \mathbf{X}^T \mathbf{T}_{\text{osc}} (\mathbf{T}_{\text{osc}}^T \mathbf{T}_{\text{osc}})^{-1} = \mathbf{X}^T \mathbf{T}_{\text{osc}}$. OSC components, containing the largest variation, are orthogonal to the concentration vector \mathbf{y} , that is, not correlated to \mathbf{y} at all, and will be removed from \mathbf{X} , $\mathbf{X}_{\text{new}} = \mathbf{X} - \mathbf{T}_{\text{osc}} \mathbf{P}_{\text{osc}}^T = \mathbf{X} - \mathbf{T}_{\text{osc}} (\mathbf{X}^T \mathbf{T}_{\text{osc}})^T = \mathbf{X} - \mathbf{T}_{\text{osc}} \mathbf{T}_{\text{osc}}^T \mathbf{X}$.

OSC has several different versions with different algorithms and often with slightly different results. These OSC algorithms have some drawbacks. The measurement error in \mathbf{y} is fully disregarded because absolute orthogonality is demanded for the OSC components in \mathbf{X} . This conflict has manifested itself in such a way that complete orthogonality is not obtained by some of the algorithms, or the final OSC component does not lie in the \mathbf{X} -space. The latter introduces new components outside the \mathbf{X} -space into the corrected matrix when the OSC component is deflated. Fearn's algorithm [61], that uses only least squares steps, does not have these problems, although it is suboptimal in describing the maximum variance of \mathbf{X} with the OSC component.

35.4.2 REGION ORTHOGONAL SIGNAL CORRECTION

In OSC the estimation of loadings \mathbf{P}_{osc} and scores \mathbf{T}_{osc} of OSC components forms its base, and the weights \mathbf{W}_{osc} are also required to calculate scores for prediction, that is, $\mathbf{t}_{\text{new}} = \mathbf{x}_{\text{new}} \mathbf{W}_{\text{osc}}$. Here, \mathbf{x}_{new} and \mathbf{t}_{new} are a spectrum (a row vector) and scores of a new sample, respectively. \mathbf{W}_{osc} , \mathbf{P}_{osc} , and \mathbf{T}_{osc} can be calculated by the training data set. The systematic variation that is unrelated to \mathbf{y} (orthogonal to \mathbf{y}) is due to a change in, for example, light scattering and differences in spectroscopic path length, and interference of other components. Clearly, the removal of OSC components concerns such systematic variation. In some cases, the source of systematic variation is very clear, for example, the background in spectra or the interference of a component is very severe. The removal of such special kinds of systematic variation is certainly demanded. As described above in normal OSC, a fixed spectral region is used for the estimation of loadings, loading weights and scores of OSC components, and the further model building with the spectra in which the OSC components are removed. That is to say, the spectra \mathbf{X} employed to estimate \mathbf{W}_{osc} , \mathbf{P}_{osc} , and \mathbf{T}_{osc} , have the same spectral region as those from which orthogonal components are removed, and those used to build a calibration model. In fact, wide region spectra measured contain information about different components, for example, a region may correspond to an absorption region for the component investigated that is used for the model building, and some regions may correspond to the background or interference components. In the spectra used, the region in the vicinity of 1440 nm shows high absorbance due to water. The spectra including this region are certainly not suitable for the model building in the determination of glucose because of the severe interference from water, but the regions including the high absorbance of water are more suitable to estimate the variation in the component of water than other spectral regions. Therefore, they are expected to become a candidate for the estimation of orthogonal components removed. Based on this basic idea, ROSC algorithm has been proposed.

In ROSC, two spectral data sets in different regions denoted by \mathbf{X}_0 and \mathbf{X} , respectively, are concerned. \mathbf{X}_0 is used to estimate weights \mathbf{W}_{rosc} and scores \mathbf{T}_{rosc} of orthogonal components, while \mathbf{X} is used in the calculation of loadings \mathbf{P}_{rosc} and the sequent removal of orthogonal components from \mathbf{X} , as well as the model building. Certainly, in the case when the same region is used for \mathbf{X}_0 and \mathbf{X} , that is, $\mathbf{X}_0 = \mathbf{X}$, ROSC and OSC are the same. The calculation procedure of ROSC is given as follows (refer to Fearn's OSC algorithm):

1. Define \mathbf{M} matrix with \mathbf{X}_0 and \mathbf{y} , $\mathbf{M} = \mathbf{I} - \mathbf{X}_0^T \mathbf{y} (\mathbf{y}^T \mathbf{X}_0 \mathbf{X}_0^T \mathbf{y})^{-1} \mathbf{y}^T \mathbf{X}_0$
2. Let $\mathbf{Z} = \mathbf{X}_0 \mathbf{M}$, and use a singular value decomposition of \mathbf{Z} to calculate singular values \mathbf{S}_0 matrix and eigenvectors \mathbf{U}_0 , $\mathbf{Z} = \mathbf{U}_0 \mathbf{S}_0 \mathbf{V}_0^T$, where the first p largest singular value λ_i ($i = 1, 2, \dots, p$) contain in the first p columns of \mathbf{S}_0 matrix, the p corresponding eigenvectors \mathbf{u}_i ($i = 1, 2, \dots, p$) contain in \mathbf{U}_0 .

3. Calculate weights of orthogonal components $\mathbf{w}_{\text{rosc},i}$ by $\mathbf{w}_{\text{rosc},i} = \mathbf{M}\mathbf{X}_0^T \mathbf{u}_i \lambda_i$, \mathbf{W}_{rosc} will be constructed by $\mathbf{W}_{\text{rosc}} = [\mathbf{w}_{\text{rosc},1}, \mathbf{w}_{\text{rosc},2}, \dots, \mathbf{w}_{\text{rosc},p}]$
4. Calculate scores of orthogonal components $\mathbf{T}_{\text{rosc}} = \mathbf{X}_0 \mathbf{W}_{\text{rosc}}$
5. Loadings of orthogonal components \mathbf{P}_{rosc} will be yielded with \mathbf{T}_{rosc} and \mathbf{X} , $\mathbf{P}_{\text{rosc}} = \mathbf{X}^T \mathbf{T}_{\text{rosc}} (\mathbf{T}_{\text{rosc}}^T \mathbf{T}_{\text{rosc}})^{-1}$
6. Remove orthogonal components from \mathbf{X} , $\mathbf{X}_{\text{new}} = \mathbf{X} - \mathbf{T}_{\text{rosc}} \mathbf{P}_{\text{rosc}}^T$
7. In the prediction for a new sample, spectra \mathbf{x}_0 and \mathbf{x} having the same spectral regions as those of \mathbf{X}_0 and \mathbf{X} , respectively, can be used, coupled with \mathbf{W}_{rosc} and \mathbf{P}_{rosc} , to calculate its scores, $\mathbf{t}_{\text{new}} = \mathbf{x}_0 \mathbf{W}_{\text{rosc}}$, and calibrate its spectra \mathbf{x} by $\mathbf{x}_{\text{new}} = \mathbf{x} - \mathbf{t}_{\text{new}} \mathbf{P}_{\text{rosc}}^T$.

35.5 REMOVAL OF INTERFERENCE SIGNALS DUE TO WATER FROM *IN VIVO* NIR SPECTRA OF BLOOD GLUCOSE BY ROSC

Du et al. [31] employed ROSC for the pretreatment of NIR spectra to remove interference signals mainly of water from the spectra in the skin spectral determination of glucose contents. Results of the application of ROSC demonstrated that ROSC-presented spectra including the whole spectral region of 1212–1889 nm or informative region of 1600–1730 nm selected by MWPLSR provide very good performance of the PLS models.

It is clearly seen from Figure 35.9 that glucose has absorption bands in the 1530–1830 nm region, while water shows high absorption signal in the wide range of 1370–1889 nm, especially in the region around 1440 nm. It is noted that there is a large amount of water in blood [31]. Therefore, water is a severe interference component for the determination of glucose. In order to search for informative and uninformative regions, MWPLSR was performed for the whole region spectra first. The calculated residue lines are shown in Figure 35.13 [31]. Two informative regions, 1240–1320 and 1600–1730 nm, can be found from Figure 35.13. The former region contains relatively weak absorption band due to glucose, but in that region there is much less interference from water (Figure 35.9). The later one is a part of the region (the 1450–1800 nm region) of glucose absorption bands. In this part the interference from water is also weak.

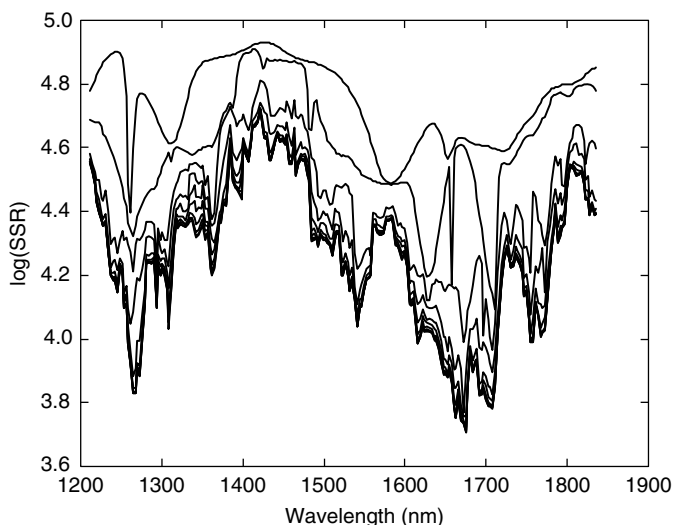


FIGURE 35.13 Residue lines obtained by MWPLSR for the original spectra. (Reproduced from Y. P. Du, Y. Z. Liang, S. Kasemsumran, K. Maruo and Y. Ozaki, *Anal. Sci.*, **20**, 1339 (2004). With permission. Copyright (2004) Elsevier.)

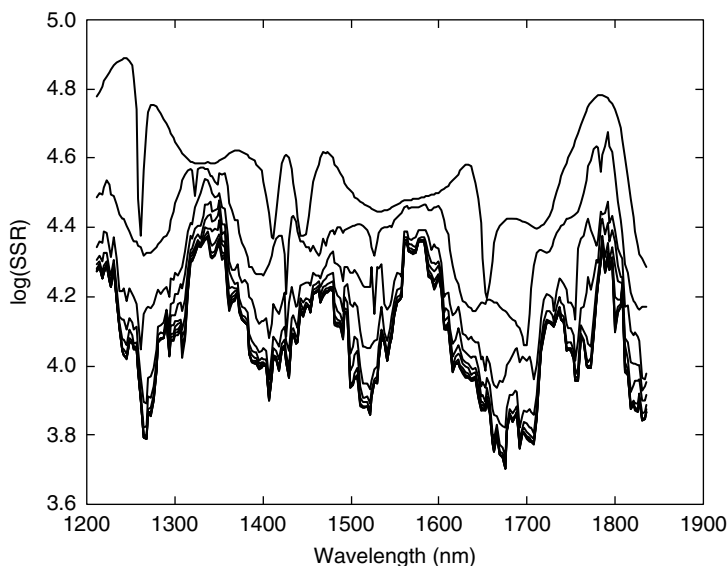


FIGURE 35.14 Residue lines obtained by MWPLSR for the OSC-pretreated spectra. (Reproduced from Y. P. Du, Y. Z. Liang, S. Kasemsunran, K. Maruo and Y. Ozaki, *Anal. Sci.*, **20**, 1339 (2004). With permission. Copyright (2004) Elsevier.)

35.5.1 SPECTRA PRETREATED BY OSC AND ROSC

OSC aims at removing strong structured variation from the whole region spectra that is orthogonal to the concentration of the interested component. OSC procedure with one orthogonal component removed was performed for the whole region spectra of the skin data set. Then, residue lines of the OSC-pretreated spectra were calculated by MWPLSR. The results are shown in Figure 35.14. Comparing with the residue lines obtained from the original spectra (Figure 35.13), the plot in Figure 35.14 provides more informative regions [31]. Informative regions of 1240–1320 and 1600–1730 nm have almost the same residue lines between Figure 35.13 and Figure 35.14 in terms of the peak positions and error levels. Figure 35.14 shows three more informative regions of 1385–1440, 1500–1550, and 1810–1889 nm but their error levels are larger than those of the ranges of 1240–1320 and 1600–1730 nm. This indicates that PLS models developed by using OSC-pretreated spectra may not yield good prediction results.

As described above, ROSC chooses a special region to estimate loading weights and scores of orthogonal components for the calculation of orthogonal components. A special region was selected according to the residue lines plot in Figure 35.13 for ROSC procedure. Note that the region of 1404–1454 nm has the highest error level in the whole region in Figure 35.13 because of the high interference of water. This region is apparently an uninformative region. Thus, the 1404–1454 nm region was selected as an uninformative region to perform ROSC because the absorption from water in this region is the highest and reflects the variation of the interference component of water (refer to Figure 35.9). With the uninformative region of 1404–1454 nm, ROSC with one orthogonal component removed was carried out. Du et al. also used the ROSC-pretreated spectra to calculate residue lines by MWPLSR, which are shown in Figure 35.15 [31]. One can see that the position of two informative regions in Figure 35.15 is similar to that of the original spectra in Figure 35.13. However, the first informative region widens its spectral range from 1240–1320 to 1240–1360 nm, and the error levels of the two informative regions, especially that of the region of 1600–1730 nm, decrease significantly. These can be explained by the decrease of water interference. Therefore, it is expected that ROSC-pretreated spectra may provide better performance of PLS models than that obtained by using the original spectra or the OSC-pretreated spectra.

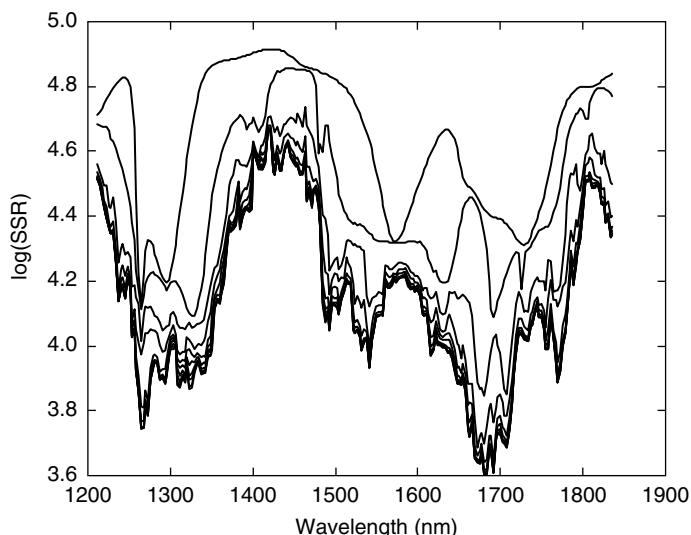


FIGURE 35.15 Residue lines obtained by MWPLSR for the ROSC-pretreated spectra. (Reproduced from Y. P. Du, Y. Z. Liang, S. Kasemsumran, K. Maruo and Y. Ozaki, *Anal. Sci.*, **20**, 1339 (2004). With permission. Copyright (2004) Elsevier.)

35.5.2 COMPARISON OF PREDICTION ABILITY BETWEEN MODELS DEVELOPED BY USING OSC- AND ROSC-PRETREATED SPECTRA

From the discussion above, the ROSC-pretreated spectra seem to be more powerful than the OSC-pretreated spectra [31]. To compare their differences, PLS models built from the OSC- and ROSC-pretreated spectra obtained by OSC and ROSC with different numbers of orthogonal components removed for the whole spectral region (1212–1889 nm) were established, and segments cross validation was used to calculate root mean square error of cross validation (RMSECV) for testing the performance of the models. Figure 35.16 compares the changes in RMSECV with the increase in the PLS components number. In Figure 35.16, the numbers marked for different lines indicate the numbers of orthogonal components removed. For OSC method (left plot in Figure 35.16), the optimal PLS components number decreases from six to two with the increase of orthogonal components removed from one to five, but RMSECV shows almost no change whose value is about 19.9 mg/dl. This result is quite similar to that reported by Fearn. Contrarily, for ROSC method (right plot in Figure 35.16), PLS components do not change (equal to six) with the increase of orthogonal components removed, but the RMSECV decreases significantly. When the number of orthogonal components removed is one, RMSECV equals to 20.2300 mg/dl, being larger than that obtained by using original spectra (19.9165 mg/dl, refer to the first row in Table 35.5). When the orthogonal components removed are more than one, RMSECV shows less value (from 18.7 to 19.1 mg/dl). Therefore, ROSC is of potential for the improvement of PLS model performance.

35.5.3 SELECTION OF SPECTRAL REGIONS TO IMPROVE PLS MODEL PERFORMANCE

Informative regions obtained by MWPLSR often show potential to improve PLS model performance, and thus, Du et al. [31] selected some informative regions to build PLS models. Informative regions for the original spectra, OSC-pretreated spectra and ROSC-pretreated spectra can be selected from

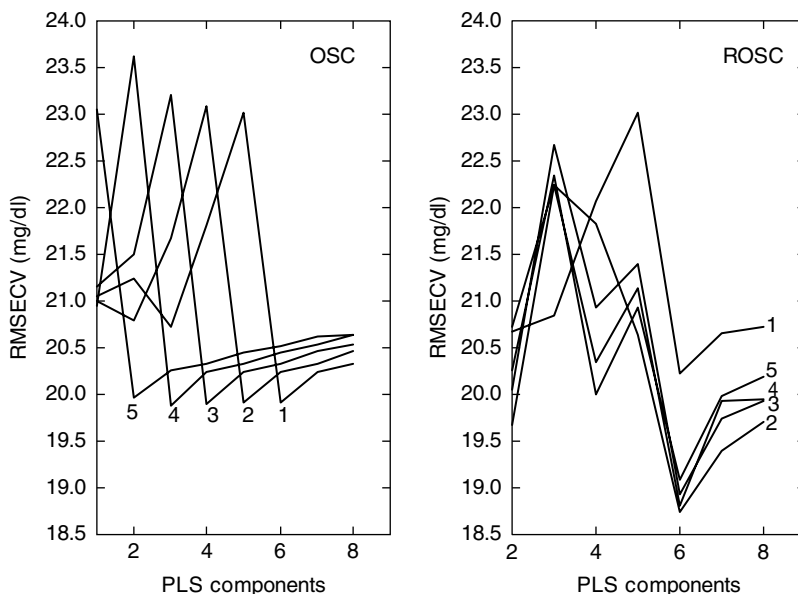


FIGURE 35.16 RMSECV of PLS models vs. PLS components with OSC- and ROSC-pretreated spectra of different orthogonal components removed in the whole region. (Reproduced from Y. P. Du, Y. Z. Liang, S. Kasemsumran, K. Maruo and Y. Ozaki, *Anal. Sci.*, **20**, 1339 (2004). With permission. Copyright (2004) Elsevier.)

TABLE 35.5

Prediction Performance of PLS Models Developed by Using Different Spectral Regions of Original Spectra

	Spectral region (nm)	PLS components number	<i>R</i>	RMSECV (mg/dl)
Whole region	1212–1889	7	.9819	19.9165
Info. region 1	1240–1320	2	.8754	23.1380
Info. region 2	1600–1730	4	.9540	17.1708

Source: K. Murayama, K. Yamada, R. Tsenkova, T. Wang and Y. Ozaki, *Near Infrared Spectrosc.*, **6**, 375 (1998).

Figure 35.13 to Figure 35.15, respectively [31]. The selected informative regions are 1240–1320 and 1600–1730 nm for the original spectra, 1240–1320 and 1600–1730 nm for the OSC-pretreated spectra, and 1240–1360 and 1600–1730 nm for the ROSC-pretreated spectra. With these selected regions and the whole region, PLS models were developed and RMSECV was calculated with the segments cross validation method for the three different kinds of spectra. Their results are listed in Table 35.5 to Table 35.7, respectively [31]. One can find from Table 35.5 that RMSECV is 19.9165 mg/dl in PLS components number of seven with the whole region of spectra, while when informative regions are used, the PLS number will reduce, and RMSECV will increase with Informative region 1 of 1240–1320 nm and decrease with Informative region 2 of 1600–1730 nm.

Normally, the use of an informative region yields a PLS model with relatively small value of RMSECV or RMSEP and small PLS dimensionality. Informative region 1 shows larger RMSECV because the absorption in this region is very low. Informative region 2 provides a good performance

TABLE 35.6

Prediction Performance of PLS Models Developed by Using Different Spectral Regions of OSC-Pretreated Spectra

	Spectral region (nm)	Orthogonal components number	PLS components number	R	RMSECV (mg/dl)
Whole region	1212–1889	1	6	.9819	19.9144
Whole region	1212–1889	2	5	.9818	19.9068
Whole region	1212–1889	3	4	.9816	19.9025
Info. region 1	1240–1320	1	1	.8751	23.1437
Info. region 2	1600–1730	1	3	.9539	17.1703
Info. region 2	1600–1730	2	2	.9536	17.1688
Info. region 2	1600–1730	3	1	.9382	17.3408

Source: Y. P. Du, Y. Z. Liang, S. Kasemsumran, K. Maruo and Y. Ozaki, *Anal. Sci.*, **20**, 1339 (2004).

TABLE 35.7

Prediction Performance of PLS Models Developed by Using Different Spectral Regions of ROSC-Pretreated Spectra

	Spectral region (nm)	Orthogonal components number	PLS components number	R	RMSECV (mg/dl)
Whole region	1212–1889	1	6	.9780	20.2300
Whole region	1212–1889	2	6	.9808	18.7406
Whole region	1212–1889	3	6	.9811	18.9237
Info. region 1	1240–1360	1	3	.9173	20.3942
Info. region 1	1240–1360	2	3	.9338	20.1154
Info. region 1	1240–1360	3	3	.9427	20.0794
Info. region 2	1600–1730	1	4	.9660	15.8911
Info. region 2	1600–1730	2	4	.9696	16.9939
Info. region 2	1600–1730	3	4	.9688	16.7170

Source: Y. P. Du, Y. Z. Liang, S. Kasemsumran, K. Maruo and Y. Ozaki, *Anal. Sci.*, **20**, 1339 (2004).

model with significant decrease of RMSECV (17.1708 mg/dl) and PLS number (4). For the OSC-pretreatment spectra (Table 35.6), when the whole region is used, as mentioned above, one can save one PLS component for each OSC component removal, but cannot improve the model performance, that is, RMSECV almost has a constant value (rows 1 to 3 in Table 35.6). Similar results can be observed in Table 35.2 (rows 5 to 7 in Table 35.6), say, OSC-pretreatment spectra in the 1600–1730 nm region cannot improve RMSECV although the number of PLS components reduces. Informative region 1 also cannot provide a satisfactory model with small value of RMSECV (refer to row 4 in Table 35.7).

Table 35.7 compares, performance of PLS models with ROSC-pretreatment spectra [31]. It is noted that RMSECV can be reduced in the case of two orthogonal components removed when the whole region is used (see rows 1 to 3 in Table 35.7). For Informative region 1 (see rows 4 to 6 in Table 35.7), although it is larger than that obtained by using the original spectra in the whole region, RMSECV (about 20.1 mg/dl) is quite smaller than that for the OSC-pretreatment spectra in the same region (Informative region 1, about 23.1 mg/dl). The increase in the number of

removed orthogonal components will reduce RMSECV slightly. It is of particular note that RMSECV significantly decreases to 15.8911 mg/dl when the ROSC-pretreatment spectra are used in Informative region 2 and one orthogonal component is subtracted (see row 7 in Table 35.7). When more than one orthogonal component are subtracted, RMSECV increases. The reason for this may be that one orthogonal component removed is enough; otherwise when more than one orthogonal component are removed, the noise signal will be concerned as the orthogonal component and result in the damnification of the model.

REFERENCES

1. H. M. Heise in *Near-Infrared Spectroscopy-Principles, Instruments, Applications*, Y. Ozaki, S. Kawata, and H.M. Heise (Eds.), Wiley-VCH, Weinheim, Germany, 2002, p. 289.
2. H. M. Heise in *Biosensors in the Body Continuous in vivo Monitoring*, D. M. Fraser (Ed.), Wiley, New York, 1997, p. 79.
3. J. A. Jaquez, J. Huss, W. McKeehan, J. M. Dimitroff and H. F. Kuppenheim, *J. Appl. Physiol.*, **8**, 297 (1955).
4. E. Peuchant, C. Salles and R. Jensen, *Anal. Chem.*, **59**, 1816 (1987).
5. J. W. Hall and A. Pollard, *Clin. Biochem.*, **26**, 483 (1993).
6. J. W. Hall and A. Pollard, *Clin. Chem.*, **38**, 1623 (1992).
7. J. W. Hall and A. Pollard in *Leaping Ahead with Near Infrared Spectroscopy*, G. D. Batten, P. C. Flinn, L. A. Welsh and A. B. Blakeney (Eds.), Royal Austral. Chem. Inst., Melbourne, 1995, p. 421.
8. G. Domjan, K. J. Kaffka, J. M. Jako and I. T. Valyi-Nagy, *J. Near Infrared Spectrosc.*, **2**, 67 (1995).
9. K. H. Hazen, M. A. Arnold and G. W. Small, *Anal. Chim. Acta*, **371**, 255 (1998).
10. K. Murayama, K. Yamada, R. Tsenkova, Y. Wang and Y. Ozaki, *Fresenius' J. Anal. Chem.*, **362**, 155 (1998).
11. S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne and Y. Ozaki, *Anal. Chim. Acta*, **512**, 223 (2004).
12. G. W. Small, M. A. Arnold and L. A. Marquardt, *Anal. Chem.*, **65**, 3279 (1993).
13. H. M. Heise, R. Marbach and A. Bittner, *J. Near Infrared Spectrosc.*, **6**, 361 (1998).
14. D. M. Haaland, M. R. Robinson, G. W. Koepp, E. V. Thomas and R. P. Eaton, *Appl. Spectrosc.*, **46**, 1575 (1992).
15. K. H. Norris and J. T. Kuenstner, in *Leaping Ahead with Near Infrared Spectroscopy*, G. D. Batten, P. C. Flinn, L. A. Welsh and A. B. Blakeney (Ed.), Royal Austral. Chem. Inst., Melbourne, 1995, p. 431.
16. K. H. Hazen, M. A. Arnold and G. W. Small, *Appl. Spectrosc.*, **48**, 477 (1994).
17. K. H. Hazen, M. A. Arnold and G. W. Small, *Appl. Spectrosc.*, **52**, 1597 (1998).
18. S. Pan, H. Chung, M. A. Arnold and G. W. Small, *Anal. Chem.*, **68**, 1124 (1996).
19. K. Murayama, K. Yamada, R. Tsenkova, Y. Wang and Y. Ozaki, *Vibrat. Spectrosc.*, **18**, 33 (1998).
20. K. Murayama, K. Yamada, R. Tsenkova, Y. Wang and Y. Ozaki, *J. Near Infrared Spectrosc.*, **6**, 375 (1998).
21. S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne and Y. Ozaki, *Analyst*, **128**, 1471 (2003).
22. M. R. Robinson, R. P. Eaton, D. M. Haaland et al., *Clin. Chem.*, **38**, 1618 (1992).
23. R. Marbach, T. H. Koschinsky, F. A. Gries and H. M. Heise, *Appl. Spectrosc.*, **47**, 875 (1993).
24. R. Marbach and H. M. Heise, *Appl. Optics.*, **34**, 610 (1995).
25. H. M. Heise, R. Marbach and T. H. Koschinsky, *Artif. Org.*, **18**, 439 (1994).
26. K. Maruo, M. Tsurugi, M. Tamura and Y. Ozaki, *Appl. Spectrosc.*, **57**, 1236 (2003).
27. M. A. Arnold and G. W. Small, *Anal. Chem.*, **62**, 1457 (1990).
28. H. M. Heise, A. Bittner and R. Marbach, *J. Near Infrared Spectrosc.*, **6**, 349 (1998).
29. S. Kasemsumran, Y. P. Du, B. Y. Li, K. Maruo and Y. Ozaki, *Analyst*, **131**, 529 (2006).
30. N. Kang, S. Kasemsumran, Y.-A. Woo, H. J. Kim and Y. Ozaki, *Chemom. Intell. Lab. Syst.*, submitted (2004).
31. Y. P. Du, Y. Z. Liang, S. Kasemsumran, K. Maruo and Y. Ozaki, *Anal. Sci.*, **20**, 1339 (2004).
32. S. Kasemsumran, Y. P. Du, K. Maruo and Y. Ozaki, *Anal. Chim. Acta*, **526**, 193 (2004).
33. K. Maruo, M. Tsurugi, J. Chin et al., *IEEE J. Sel. Top. Quant.*, **9**, 322 (2003).

34. J. J. Burmeister and M. A. Arnold, *Clin. Chem.*, **45**, 1621 (1999).
35. M. A. Arnold, J. J. Burmeister and G. W. Small, *Anal. Chem.*, **70**, 1773 (1998).
36. S. F. Malin, T. L. Ruchti, T. B. Blank, S. N. Thennadil and S. L. Monfre, *Clin. Chem.*, **45**, 1651 (1999).
37. C. Fischbacher, K. U. Jagemann, K. Danzer, U. A. Müller, L. Papenkordt and J. Schüler, *Fresenius J. Anal. Chem.*, **359**, 78 (1997).
38. B. G. M. Vandegiste, D. L. Massart, L. M. C. Buydens, S. de Jong, P. L. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, Amsterdam, 1998.
39. H. Martens and T. Næs, in *Multivariate Calibration*, John Wiley: Chichester, 1989.
40. J. H. Jiang, R. J. Berry, H. W. Siesler and Y. Ozaki, *Anal. Chem.*, **74**, 3555 (2002).
41. Y. P. Du, Y. Z. Liang, J. H. Jiang, R. J. Berry and Y. Ozaki, *Anal. Chim. Acta*, **501**, 183 (2004).
42. S. Kasemsumran, Y. P. Du, K. Maruo and Y. Ozaki, *Chemom. Intell. Lab. Syst.*, **82**, 97 (2006).
43. A. Lober and B. R. Kowalski, *J. Chemom.*, **2**, 67 (1988).
44. L. Xu and I. Schechter, *Anal. Chem.*, **68**, 2392 (1996).
45. C. H. Spiegelman, M. J. McShane, M. J. Goetz, M. Motamedi, Q. L. Yue and G. L. Cote, *Anal. Chem.*, **70**, 35 (1998).
46. J. H. Kalivas, N. Roberts and J. M. Sutter, *Anal. Chem.*, **61**, 2024 (1989).
47. D. J. Rimbaud, B. Walczak, D. L. Massart, I. R. Last and K. A. Prebble, *Anal. Chim. Acta*, **304**, 185 (1995).
48. P. J. Brown, *J. Chemom.*, **6**, 151 (1992).
49. J. H. Kalivas, N. Roberts and J. M. Sutter, *Anal. Chem.*, **61**, 2024 (1989).
50. Y. Z. Liang, Y. L. Xie and R. Q. Yu, *Anal. Chim. Acta*, **222**, 347 (1989).
51. U. Horchner and J. H. Kalivas, *Anal. Chim. Acta*, **311**, 1 (1995).
52. C. B. Lucasius and G. Kateman, *Trends Anal. Chem.*, **10**, 254 (1991).
53. L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck and S. B. Engelsen, *Appl. Spectrosc.*, **54**, 413 (2000).
54. L. Munck, J. P. Nielsen, B. Møller et al., *Anal. Chim. Acta*, **446**, 171 (2001).
55. S. Wold, *Technometrics*, **20**, 397 (1978).
56. H. A. Martens and P. Dardenne, *Chemom. Intell. Lab. Syst.*, **44**, 91 (1998).
57. J. Shao, *J. Am. Stat. Assoc.*, **88**, 486 (1993).
58. Q. S. Xu and Y. Z. Liang, *Chemom. Intell. Lab. Syst.*, **56**, 1 (2001).
59. S. Wold, H. Antti, F. Lindgren and J. Ohman, *Chemom. Intell. Lab. Syst.*, **44**, 175 (1998).
60. J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg and S. Wold, *Chemom. Intell. Lab. Syst.*, **44**, 229 (1998).
61. T. Fearn, *Chemom. Intell. Lab. Syst.*, **50**, 47 (2000).
62. B. M. Wise and N. B. Gallagher, <http://www.eigenvector.com/MATLAB/OSC.html>

36 Plastics Analysis at Two National Laboratories

CONTENTS

Part A Resin Identification Using Near-Infrared Spectroscopy and Neural Networks...	699
36A.1 Introduction.....	699
36A.2 Theory	701
36A.2.1 Artificial Neural Networks	701
36A.3 Experimental	701
36A.3.1 Field Instrumentation	703
36A.4 Results and Discussion	704
36A.4.1 Spectroscopy	704
36A.4.2 Artificial Neural Networks	705
36A.4.3 Field Instrumentation	706
36A.5 Conclusions	707
Acknowledgments	708
References	708
Part B Characterization of Plastic and Rubber Waste in a Hot Glovebox.....	710
36B.1 Introduction.....	710
36B.2 Instrumentation.....	711
36B.3 Results	713
36B.4 Discussion	715
36B.5 Conclusion.....	715
References	716

This chapter consists of two parts, both dealing with the use of near-infrared (NIR) spectroscopy to identify plastic materials at National Laboratories. Part A describes work done at Sandia National Laboratory and Part B describes work done at Los Alamos National Laboratory.

PART A

RESIN IDENTIFICATION USING NEAR-INFRARED SPECTROSCOPY AND NEURAL NETWORKS

*M. Kathleen Alam, Suzanne Stanton, and
Gregory A. Hebner*

36A.1 INTRODUCTION

In Europe, quotas are in place that will mandate that 15% of the waste headed to landfills be recovered in recycling programs. While the U.S. government has yet to implement such

a strong policy, some individual states have begun to require some form of recycling, for instance, implementing regulations requiring products to contain a certain percentage of recycled material.

Plastic waste recycling is difficult due to the costs associated with it. In the United States and abroad, recovered plastic wastes have been used to replace virgin materials [1] as filler materials for wood and concrete composites [2–4], and as sources for feedstock recycling and energy recovery [5]. While the degree of resin purity required for each end use is different, in all cases minimal separation of the resins is desired [6]. The most preferred route for recycled plastic materials is the replacement of virgin resin in manufacturing. However, the costs of collecting, sorting, and processing of polymer waste is more expensive than manufacturing items from virgin materials [7,8]. Sorting is a key step in the recovery process, as the quality, strength, and subsequently, the value of the resulting product increases the purer the incoming resin stream.

Common methods for sorting waste plastics are hydrocycloning, which separates the polymer species based on their densities, and labor-intensive hand sorting, both of which are error prone. Newer technologies for separating polymers from other materials include electrostatic separation, and pneumatic table separation [9]. For separating individual polymers, more advanced technologies have been developed for sorting waste plastics based on spectroscopic measurements. X-ray fluorescence has been used to separate polyvinyl chloride (PVC) from other plastics via x-ray fluorescence spectroscopy [10]. Several visible, infrared (IR), and Raman methods have been developed to sort waste plastics by color and resin type [10–19].

A potential resin sorting system must have high throughput. Industry requires that a minimum of five items be sorted per second with a high degree of accuracy. IR methods have the capability to meet or exceed these requirements due to the fast spectral acquisition of IR instrumentation, and the unique absorption patterns of polymer resins. Near-infrared reflectance spectroscopy (NIRS) is especially well suited as a possible resin sensor, due to the relative ease of sample presentation, and the relative insensitivity to minor components within the resin matrix (i.e., pigments, additives). Because the current cost of recycled resin is low and the profit margins are small, the NIRS sensor must be inexpensive so that the investment cost can be quickly recuperated. The hostile environment of most recycling plants (e.g., dirt, water, temperature extremes) necessitates a rugged system as well. If repairs are required, the system should be modular to reduce the need for highly trained technicians.

An integral part of a NIRS resin sensor must be a robust classification method. Univariate classification of NIRS spectra collected from resin is difficult, if not impossible, due to the overlapping nature of NIR spectral bands, and the variation in scattering due to the variability in sample presentation. Multivariate preprocessing and analysis methods are capable of providing the increased precision and accuracy needed. Multivariate pattern recognition methods such as Mahalanobis distance, K-Nearest Neighbor (KNN), and Soft Independent Modeling of Class Analogies (SIMCA) have shown the ability to provide accurate classification of spectral data [20–25].

Neural network technology has also shown promise for spectral data analysis [26–28]. Originally conceived as models for human learning and cognition, artificial neural networks (ANNs) have seen wide use in the fields of speech and image recognition [29], as well as in sound recognition [30], and process control [31,32]. ANNs have the ability to filter noise, as do many of the multivariate pattern recognition techniques, and they have the ability to model offsets and slopes in the spectral data. They may also be configured to check incoming data for deviations, and if needed, update the model off line.

The combination of NIRS and ANNs is presented here as a potential sensor for separating waste plastics according to resin type. Laboratory results indicate good separation of five of the six categories of resins commonly found in consumer items. Results using a field prototype are also presented.

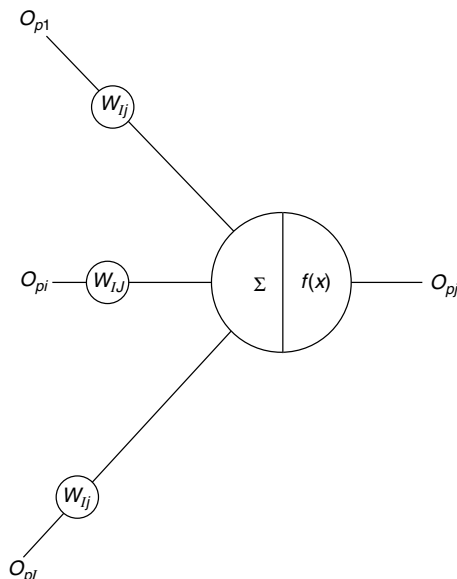


FIGURE 36.1 Schematic representation of a neural network node, j . The inputs into node j are represented by O_{Lj} , O_{ij} , and O_{jj} , where i represents an arbitrary transmitting node and L is the total number of transmitting nodes. The weights, W_{Lj} , through W_{ij} , are applied to each input value. Each weight is determined iteratively, as the network trains. Weighted inputs are summed, then transformed according to the function $f(x)$.

36A.2 THEORY

36A.2.1 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are composed of elements called nodes that process a series of weighted inputs (Figure 36.1). The input signals, O_{pi} , are weighted by scalars, W_{ij} , which are determined iteratively during the training process. The weighted inputs are summed, then transformed using a linear or nonlinear transfer function, $f(x)$. A common transfer function is the sigmoid:

$$f(x) = [1 + e^{(-x/\theta)}]^{-1} \quad (36.1)$$

The value x in Equation (36.1) refers to the weighted summation of input signals. The parameter θ refers to the gain, and is usually set by the user. The gain modifies the shape of the sigmoid curve and serves to increase or decrease the amount of nonlinearity in the model. Because the results of this transfer function are limited to values between 0 and 1, outputs of the individual nodes are subsequently limited to this range as well. Nodes are grouped together into layers, and a network can have several layers. A common architecture calls for the network to be fully connected and feed forward, that is, all nodes on one layer send their output signals forward to all nodes on the following layer (as an example, see Figure 36.2). Given training data, a network alters its weights until the “correct” answers are achieved. In the present application, a multilayer back-propagation method is used to adjust the weights. Back-propagation involves using the errors at each output as factors in adjusting the weights, thus “back-propagating” the error through the network [33]. Weights are altered until the root mean square (r.m.s.) error of the network is below a minimum threshold.

36A.3 EXPERIMENTAL

Household plastic samples were collected from various sources. The resin type of each item collected was identified by SPI (Society of Plastic Industries) codes printed on the items. The codes consist of

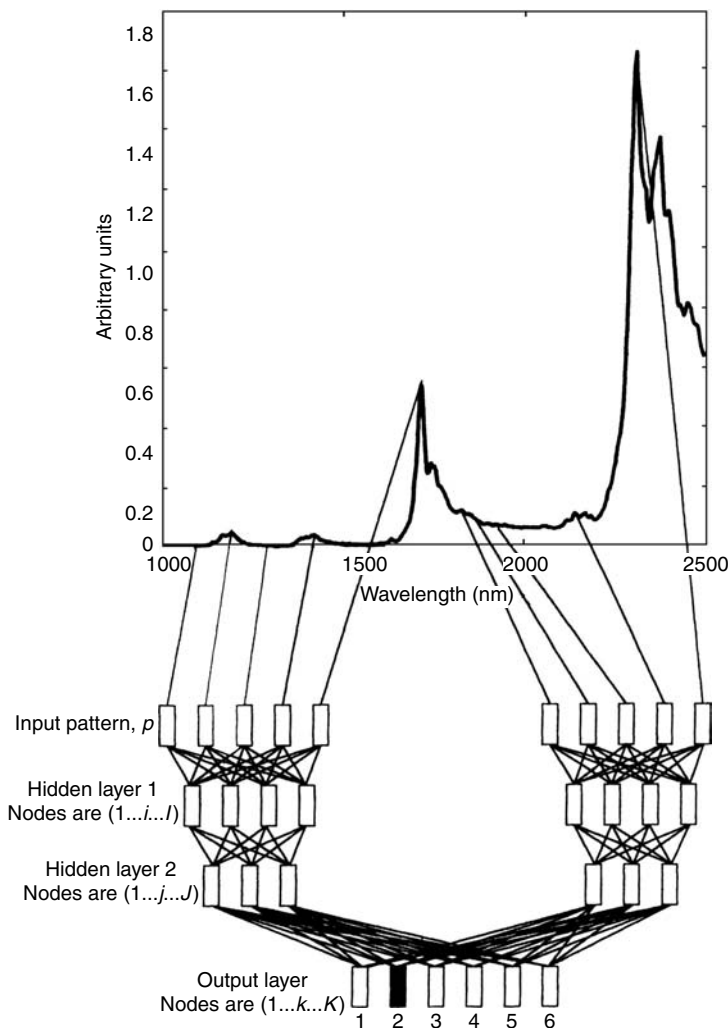


FIGURE 36.2 Schematic representation of a multilayer network. For waste plastic classification, spectral reflectance values provide inputs (IN) into the network. Two hidden layers HID1 and HID2 are used, composed of a variable number of nodes. The outputs of the network (OUT), indicate the resin from which the spectrum was collected.

a number, 1 to 7, inside a triangle. The numbers refer to the resin type: 1-polyethylene terephthalate (PET), 2-high density polyethylene (HDPE), 3-polyvinyl chloride (PVC), 4-low density polyethylene (LDPE), 5-polypropylene (PP), and 6-polystyrene (PS). Items with code 7, which refers to co-layered items or other resin materials, were not included in the study. Training and test sets were developed from the collected samples. The training set consisted of 80 items, and the test set consisted of 63 items. Coupon samples, approximately $2'' \times 2''$ in size, were collected from each item. Adhesives, grime, and other impurities present on the coupons were not removed. Paper labels were removed, or avoided when coupons were collected. Various angles, shapes, and colors were purposely chosen for coupons to mimic the differing attributes of the larger samples.

NIR reflectance spectra were collected using a Laser Precision PCM 4000 Fourier transform near-infrared (FT-NIR) spectrometer, equipped with CaF beam splitters and a thermoelectrically cooled PbSe detector. An Axiom diffuse/specular reflectance attachment, set at 15°C , was used to collect the reflectance spectrum from each sample coupon. Each sample spectrum was the result of a 5 scan

TABLE 36.1
Samples and Spectra within Training and Test Sets

Training set		
Category	Number of spectra	Number of samples
PET	78	23
HDPE	79	19
PVC	102	8
LDPE	86	11
PP	73	9
PS	77	10
Totals	494	80
Test set		
Category	Number of spectra	Number of samples
PET	11	11
HDPE	10	10
PVC	10	10
LDPE	10	10
PP	13	13
PS	9	9
Totals	63	63

average, collected at 4 cm^{-1} resolution, covering the spectral range 1000 to 2500 nm, for a total of 1556 digitized points. Reflected sample intensities were ratioed to those from a standard ceramic disk (Coors Ceramic Company, Golden, CO). $\log(1/R)$ values were calculated and used for analysis. The entire sample spectrum, including scan averaging, fast Fourier transform (FFT), and the calculation of the $\log(1/R)$ values, took approximately 10 s. Spectra were collected at various locations on each sample coupon in the training set, for a total of 494 training spectra. A single spectrum was collected for each test set coupon for a total of 63 test spectra. The number of spectra collected for each resin category, in both the training set and test sets is listed in Table 36.1. Preprocessing of spectral data was performed using routines written in MATLAB (The Mathworks, Natick, MA), operating on either a PC-486 or a SUN Workstation.

ANNs, using multilayer back-propagation architectures were created using the software package NeuralWorks II Professional Software Package (NeuralWare, Inc., Pittsburgh, PA), running on a SUN Workstation. The number of input nodes to the network was variable. Both single hidden layer and two-layer networks were trained and tested. Networks with two hidden layers (HID) were found to perform optimally. Both six- and five-output networks were evaluated. For the six-output network, each output corresponded to one of the six distinct polymer types, while for the five-output network, HDPE and LDPE resin were combined into a single category. The learning rate during standard back-propagation training was set at 0.9, momentum was set at 0.0, and the gain of the sigmoid transfer function was set at 1.0. The learning rate and momentum are used in calculating the changes in the weights. The final convergence criterion was $\pm 0.5\%$ r.m.s. error.

36A.3.1 FIELD INSTRUMENTATION

The prototype instrument was based upon a spinning circular variable wavelength filter (CVF) and utilized a low-noise InAs detector. A CVF is a time-dependent wavelength selector that passes a narrow spectral band. The spectral region of the band depends on the angular position of the filter.

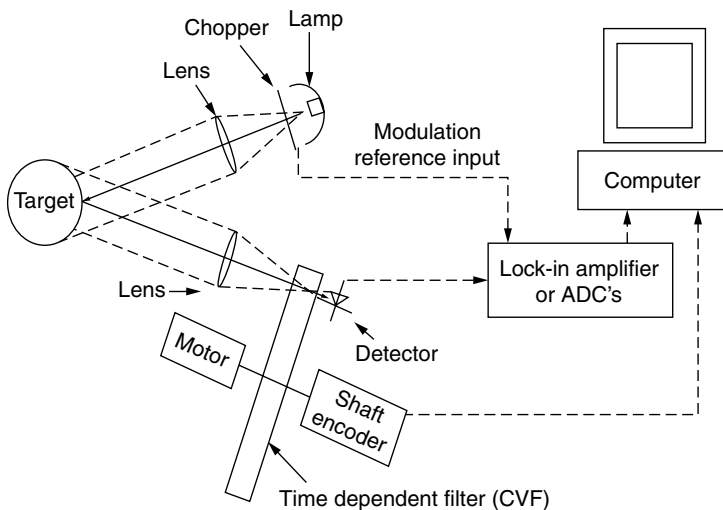


FIGURE 36.3 Schematic diagram of CVF-based instrument used for resin identification.

Light from a tungsten halogen lamp was modulated by a rotating chopping wheel. Lenses collected and collimated the modulated light, sending it to the plastic target. Reflected light from the plastic target was collected and focused through the CVF onto a low-noise IR detector (cooled InAs photodiode). A motor coupled to the filter shaft rotated the filter. An encoder attached to the motor indicated the position of the shaft as a function of time.

The output from the detector was measured using a lock-in amplifier. The lock-in amplifier amplified frequencies that were within a narrow frequency band as determined by a reference frequency. In this case, the reference frequency was the frequency of the modulated light from the lamp. All background noise at frequencies not within the pass band of the lock-in amplifier was rejected.

An IBM-compatible computer was used to assimilate the data. The detector output provided the reflected signal as a function of time while the encoder on the shaft of the CVF provided the filter wavelength as a function of time. The computer processed the information to give reflectance as a function of wavelength. A schematic of the device is shown in Figure 36.3.

36A.4 RESULTS AND DISCUSSION

36A.4.1 SPECTROSCOPY

NIR reflectance spectra of the six types of recyclable plastics collected for this study are shown in Figure 36.4. Bands seen are primarily those arising from overtones and combinations of C—H and O—H stretching and bending vibrations. Combination bands of C—H are found between 2100 and 2500 nm while overtones of aromatic, methyl, and methylene C—H stretches are found between 1600 and 1800 nm and between 1150 and 1230 nm.

Within a resin category, the location of the spectral bands were constant. However, offsets, slopes, and the maximum optical density (OD) varied considerably due to the variability in each sample's color, shape, thickness, and orientation in the spectrometer. Figure 36.5 shows several spectra of polystyrene, collected from various samples. Baseline offsets are present with typical variations on the order of ± 0.25 OD and a maximum deviation of 1.7 OD for a black-pigmented polystyrene sample. Sloping offsets are also apparent in several of the spectra. While preprocessing can eliminate some of the deviations, others must be incorporated into the model.

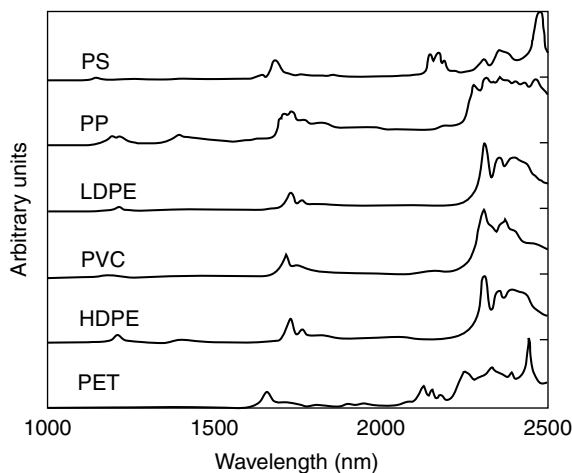


FIGURE 36.4 Reflectance spectra collected from post-consumer plastics for the NIR spectra region 1,000 to 2,500 nm ($10,000$ to $4,000\text{ cm}^{-1}$), shown as $\log(1/R)$. Spectra have been offset for clarity. Each spectrum is the result of a 5 scan average.

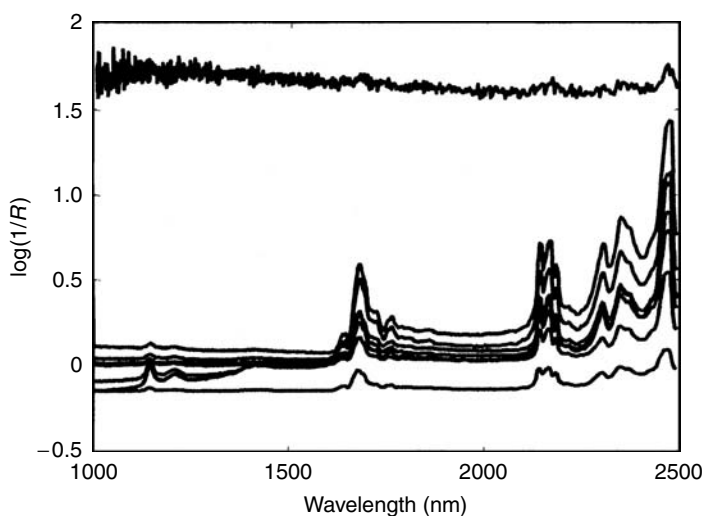


FIGURE 36.5 Subset of reflectance spectra collected from PS samples $\log(1/R)$. Note the variation in slope intensity, and baseline offset. The spectrum with a 1.7 OD offset is from a black pigmented PS sample.

36A.4.2 ARTIFICIAL NEURAL NETWORKS

Optimization of the networks is described elsewhere [18]. Preprocessing effects, number of hidden layers, number of nodes within a layer, and types of transfer functions were evaluated. Every 18th point of the 1556-point spectrum was used for input data, for a total of 87 points. The optimal network for the six-output application was determined to be a three-layer network, consisting of the 87 inputs, 15 nodes on the first hidden layer, 11 nodes on the second hidden layer, and the 6-node output layer (87-15-11-6 network). The optimal preprocessing method for the input spectra consisted of subtracting the mean spectral intensity of each sampled from itself and then dividing the result by the standard deviation of the spectral intensities in

TABLE 36.2
Neural Network Predicted Value of Training Data

		PET	HDPE	PS	LDPE	PP	PS
Known	PET	78	0	0	0	0	
	HDPE	0	71	0	3	0	0
Resin Type	PVC	0	0	102	0	0	0
	LDPE	0	8	0	83	0	0
	PP	0	0	0	0	73	0
	PS	0	0	0	0	0	77

TABLE 36.3
Neural Network Predicted Value of Test Data

		PET	HDPE	PS	LDPE	PP	PS
Known	PET	11	0	0	0	0	
	HDPE	0	8	0	2	0	0
Resin Type	PVC	0	0	10	0	0	0
	LDPE	0	2	0	8	0	0
	PP	0	0	0	0	13	0
	PS	0	0	0	0	0	9

the sample:

$$x_{i'j} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (36.2)$$

where x_{ij} is the reflectance value at the i th wavelength of the j th sample, \bar{x}_j is the mean of all the reflectance values in the j th sample, and s_j is the standard deviation of all the reflectance values in the j th sample.

Results for the training and test sets using the described 6-output network are shown in Table 36.2 and Table 36.3. Results are shown as the number of samples identified within the resin category. The tables show that using the 87-15-11-6 network with the normalized data, the only errors noted were in the distinction of LDPE and HDPE plastics. This is not surprising, because the molecular structure of LDPE and HDPE is so similar. Reconstructing the network with 5 outputs (combining LDPE and HDPE), gave 100% training and test classification.

36A.4.3 FIELD INSTRUMENTATION

The first prototype developed, described previously, is a high-speed scanning NIR spectrometer that is capable of producing a reflectance spectrum from 1300 to 2400 nm in 75 ms. A schematic of the system is shown in Figure 36.3. Reflectance spectra obtained using the prototype device are shown in Figure 36.6. Figure 36.6a–f correspond to the plastic resins PET, HDPE, PVC, LDPE, PP, and PS respectively. Using the current system, we have demonstrated a spectral acquisition time of 75 ms or 13.3 spectra/s. The background was not removed from the data prior to data analysis.

Despite the significant background component contained in the data, the trained neural network was capable of predicting the six standard resin types with accuracy greater than 96% without background subtraction. Again, errors in prediction were those involving LDPE and HDPE. The

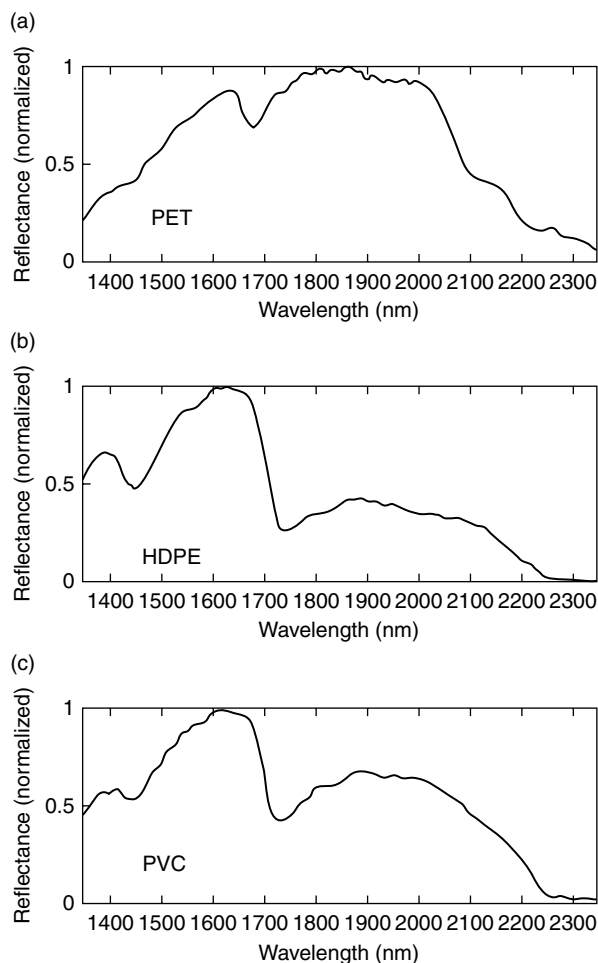


FIGURE 36.6 Single-beam spectra collected from post-consumer plastics for the NIR spectral region 1300 to 2400 nm, shown as Reflectance. Each spectrum was collected in 75 ms.

background component (predominantly due to the spectral features of the quartz halogen IR lamp) does not represent a limit to this implementation. The spectra produced by the CVF have lower resolution than the FTIR spectra, yet initial studies indicate that the neural network is capable of identifying the six standard plastic resins with a very high degree of accuracy.

Although the 75 ms acquisition time is a significant improvement over laboratory-based collection times, some conveyer speeds may require a faster collect time. A second prototype, not detailed here, incorporated acousto optical tunable filter (AOTF) technology that allowed faster, and more selective data collection.

36A.5 CONCLUSIONS

NIRS combined with neural network technology has shown the ability to provide a very capable sensor for sorting waste plastics. In both laboratory and field testing, greater than 95% accuracy was achieved on test sets, with errors limited to distinguishing between HDPE and LDPE. Relatively inexpensive implementations of the technology may provide a much-needed solution for the sorting needs of the recycling industry.

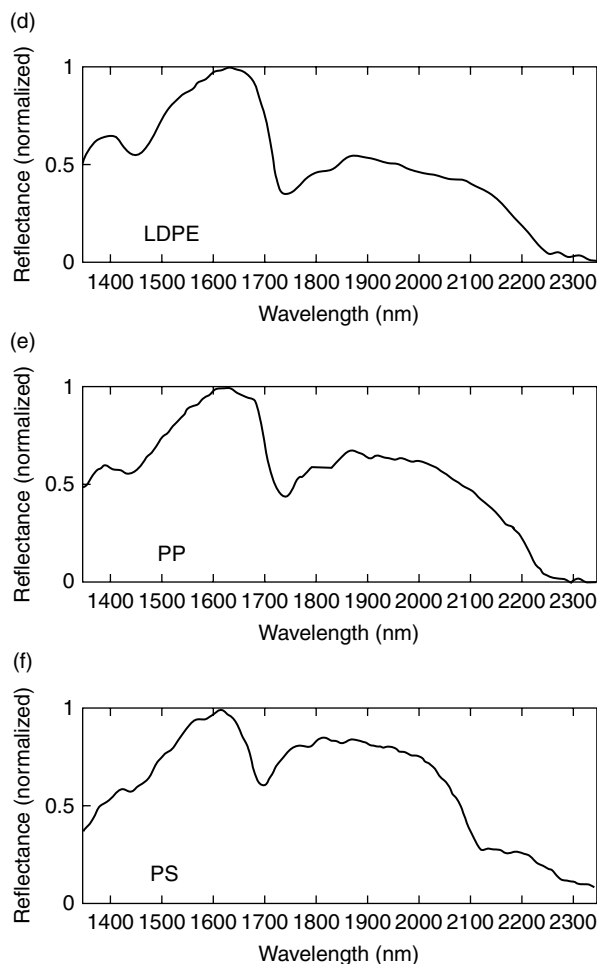


FIGURE 36.6 (Continued.)

ACKNOWLEDGMENTS

This work was performed at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000. Thanks go to Chris Coil, Pat Nelson, Brigetta Bauer, Jeanne Bando, Kevin Killeen, and Regan Montaño for aiding in data collection and analysis. Work by Tom Hamilton with the field prototype was most appreciated. Finally, we thank Sandia's Laboratory Directed Research and Development Program for supporting this research.

REFERENCES

1. S. Karlsson and A.-C. Albertsson, *Recycling of Cheap Packaging Waste Versus Expensive Engineering Materials*, J. Kahovec (ed.), 38th Microsymposium on Recycling of Polymers, Prague, Czech Republic, Wiley, VCH (1997).
2. C. Zandonella, *New Scientist*, 162: 26 (1999).
3. F. March, *Sea Technology*, 37: 35 (1996).
4. K. Rebeiz, S. Yang, and D. Fowler, *ACI Materials Journal*, 91: 313 (1994).

5. J. Brandrup, *Ecological and Economical Aspects of Polymer Recycling*, J. Kahovec (ed.), 38th Microsymposium on Recycling of Polymers, Prague, Czech Republic, Wiley, VCH (1997).
6. N. M. Bikales, *Final Report of the IUPAC Working Party on Recycling of Polymers*, J. Kahovec (ed.), 38th Microsymposium on Recycling of Polymers, Prague, Czech Republic, Wiley, VCH (1997).
7. C. Hadjilambrinos, *Environmental Conservation*, 24: 298 (1996).
8. K. R. Kreisher, et al., *Modern Plastics*, April: 1 (1990).
9. W. Michaeli and K. Breyer, *Polymer Recycling — Status and Perspectives*, J. Kahovec (ed.), 38th Microsymposium on Recycling of Polymers, Prague, Czech Republic, Wiley, VCH (1997).
10. J. A. Schut, *Plastics Technology*, 38: 15 (1992).
11. C. T. McCourt, et al., Computer-Controlled System and Method for Sorting Plastic Items, U.S. Patent 5,150,307, 1992.
12. H. D. Ruhl, Jr. and K. R. Beebe, Method for Sorting Used Plastic Containers and the Like, U.S. Patent 5,134,291, 1992.
13. A. Murase and N. Sato, *Applied Spectroscopy*, 53: 745 (1999).
14. D. M. Scott, *Measurement Science and Technology*, 6: 156 (1995).
15. C. E. Tinney, B. Thomson, and E. Johnson, *Spectroscopy*, 11: 50 (1996).
16. K. J. Allen and R. G. Rodriguez, *Applied Spectroscopy*, 53: 672 (1999).
17. D. Wienke, et al., *Analytical Chimica Acta*, 317: 1 (1995).
18. M. K. Alam, S. L. Stanton, and G. A. Hebner, *Spectroscopy*, 9: 30 (1994).
19. J. Florestan, et al., *Resources Conservation and Recycling*, 10: 67 (1994).
20. H. L. Mark and D. Tunnel, *Analytical Chemistry*, 57: 1449 (1985).
21. H. Mark, *Analytical Chemistry*, 59: 790 (1987).
22. B. R. Kowalski and C. F. Bender, *Analytical Chemistry*, 44: 1405 (1972).
23. B. R. Kowalski and C. F. Bender, *Journal of the American Chemical Society*, 94: 5632 (1972).
24. P. J. Gemperline, L. D. Webber, and F. O. Cox, *Analytical Chemistry*, 61: 138 (1989).
25. S. Wold and M. Sjostrom, SIMCA: A Method for Analysing Chemical Data in Terms of Similarity and Analogy, in *Symposium on Chemometrics, Theory and Application*, B. R. Kowalski (ed.), Washington, DC: American Chemical Society, 1977, vol. 52, p. 243.
26. R. Feldhoff, D. Wienke, K. Cammann, and H. Fuchs, *Applied Spectroscopy*, 51: 362 (1997).
27. H. Yang, J. Jegla, and P. Griffiths, *Fresenius Journal of Analytical Chemistry*, 362: 25 (1998).
28. M. E. Munk, M. S. Madison, and E. W. Robb, *Journal of Chemical Information and Computer Sciences*, 36: 231 (1996).
29. R. P. Lippmann, *IEEE ASSP*, 4: 4 (1987).
30. R. P. Gorman and R. J. Sejnowski, Sonar Classifier, in *DARPA Neural Network Study, Oct. 1987 to Feb. 1988*, AFCEA International Press, Fairfax, VA, 1988, p. 421.
31. R. S. Sutton, GTE Process Monitor, in *DARPA Neural Network Study, Oct. 1987 to Feb. 1988*, AFCEA International Press, Fairfax, VA, 1988, p. 411.
32. D. Nguyen and B. Windrow, The Truck Backer-Upper: An Example of Self-Learning in Neural Networks, in *International Joint Conference on Neural Networks, IJCNN*, IEEE, Washington, DC, 1989, vol. 2, p. 357.
33. J. R. Long, V. G. Gregoriou, and P. J. Gemperline, *Analytical Chemistry*, 62: 1791 (1990).

PART B**CHARACTERIZATION OF PLASTIC AND RUBBER
WASTE IN A HOT GLOVEBOX**

Donald A. Burns

36B.1 INTRODUCTION

Waste has accumulated ever since studies with radioactive materials were first undertaken. Typical waste consists of plastic, rubber, glass, and metal equipment as well as clothing, other cloth, and paper products. At Los Alamos National Laboratory, for example, about 40,000 drums of waste contaminated with Pu and Am radionuclides were put into temporary storage awaiting transfer to a permanent storage facility.

Government regulations required that the contents of these drums be identified prior to any transfer to another site. (Editors note: that site, now in use, is the Waste Isolation Pilot Plant (WIPP) near Carlsbad in southeastern New Mexico.) Accordingly, a project was initiated in which representative drums were opened in a total containment facility — a special room where each individual item of waste could be safely analyzed. The room was essentially a huge glovebox fitted with several windows and sealed-in gloves on two of its walls. Inside, a remotely controlled crane could open a drum and pour its contents onto a movable table so that several persons could handle each item (i.e., weigh, measure, identify, photograph, and analyze) prior to returning it to its drum. The diagram in Figure 36.7 is a top view of the room.

NIRS was chosen as the analytical method of choice for the plastics and rubber materials because the literature supported the use of this nondestructive technique when C—H, N—H, or O—H chemical bonds were involved. Plastics are polymers in which the repeating units vary considerably. For example, polyethylenes have a repeating unit consisting of only CH₂ units. If one of the hydrogens is changed to a methyl group (Me), it becomes PP. If a phenolic group (Ph) is attached, it becomes PS. Substituting a chlorine atom (Cl) produces PVC whereas substituting a fluorine atom for each of the four hydrogens produces Teflon. Modifying the carbon backbone to include nitrogen and oxygen creates the polyurethane family, which can be either

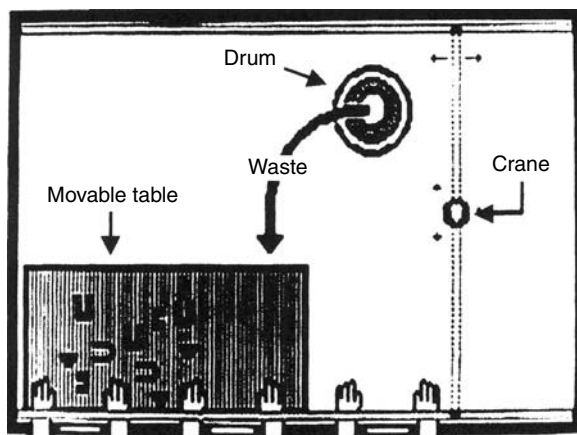


FIGURE 36.7 Top view of glovebox room showing remote-controlled table, overhead crane for dumping contents of barrel onto table, and three of several sets of built-in gloves.

ether- or ester-based (i.e., $-\text{C}(\text{OH})-\text{O}-$ or $-\text{C}(\text{=O})-\text{O}-$). Some of these structures are shown as follows:

$-\text{[CH}_2-\text{CH}_2\text{]}_n-$	$-\text{[CMeH-CH}_2\text{]}_n-$	$-\text{[CPhH-CH}_2\text{]}_n-$	$-\text{[CClH-CH}_2\text{]}_n-$	$-\text{[NH-CHO]}_n-$
polyethylene	polypropylene	polystyrene	polyvinylchloride	polyurethane

This study involved the examination of fifteen 55-gallon drums containing heterogeneous waste (combustibles). One hundred eighty-three samples were subjected to qualitative analysis by NIRS and over 82% were positively identified. The rest of Part B describes the instrumentation employed, how the analyses were performed, the special sample holder, and the results/discussion.

36B.2 INSTRUMENTATION

A top-of-the-line scanning monochromator (FOSS/NIRSystems Model 6500 with a remote reflectance probe at the end of a five-foot fiber optic cable) was used with the vendor's software (NSAS and IQ²). It was mounted on a bracket above a window outside the large enclosure that would receive the drums and contain their contents during opening, identification, and repackaging. Cables connected the monochromator to both the sensor box containing the detector probe and a computer on a rolling cart (Figure 36.8).

To avoid contaminating the instrument, a modification had to be made to one of the glovebox windows (typically made of safety glass). Because safety glass is fabricated with a layer of plastic between two pieces of glass, the sandwiched plastic would create an analytical dilemma. Accordingly, the modification consisted of removing a 3 in. diameter circle in the window and replacing it with a similar size/shape piece of quartz, thereby avoiding the need to look through one piece of plastic while attempting to identify other plastics. Figure 36.9 shows how the quartz window simplified the spectrum of the background.

So information about the sample could be obtained in a noninvasive manner, a means was needed to hold samples of varying sizes and shapes in such a position within the glovebox so that the remote reflectance probe outside the glovebox would always be in the correct alignment. Reference to Figure 36.10 will simplify the description that follows. A hollow Teflon piston about 3" square \times 5" long contained a spring-loaded loop of Teflon wire, part of which rested on the flat surface of one end (top left). The piston fit snugly inside a square inner tube whose inside surface was lined with black felt for a light-tight seal (top right). A sample, up to the size of a man's fist, was held on the

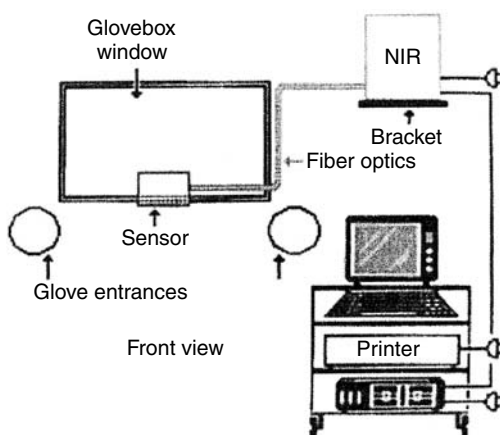


FIGURE 36.8 Diagram of major components outside the glovebox.

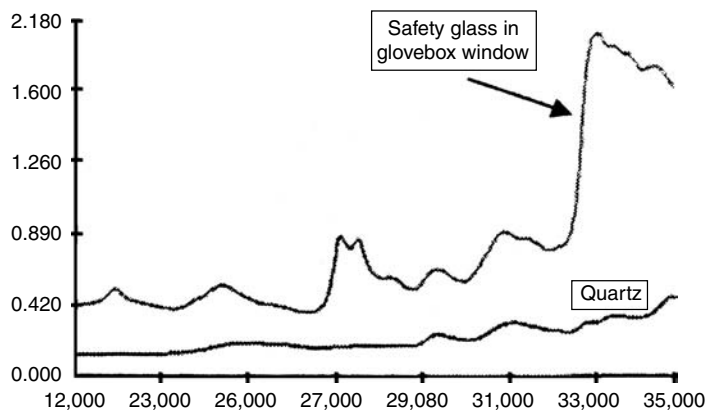


FIGURE 36.9 NIR tracings with and without quartz insert in glovebox window.

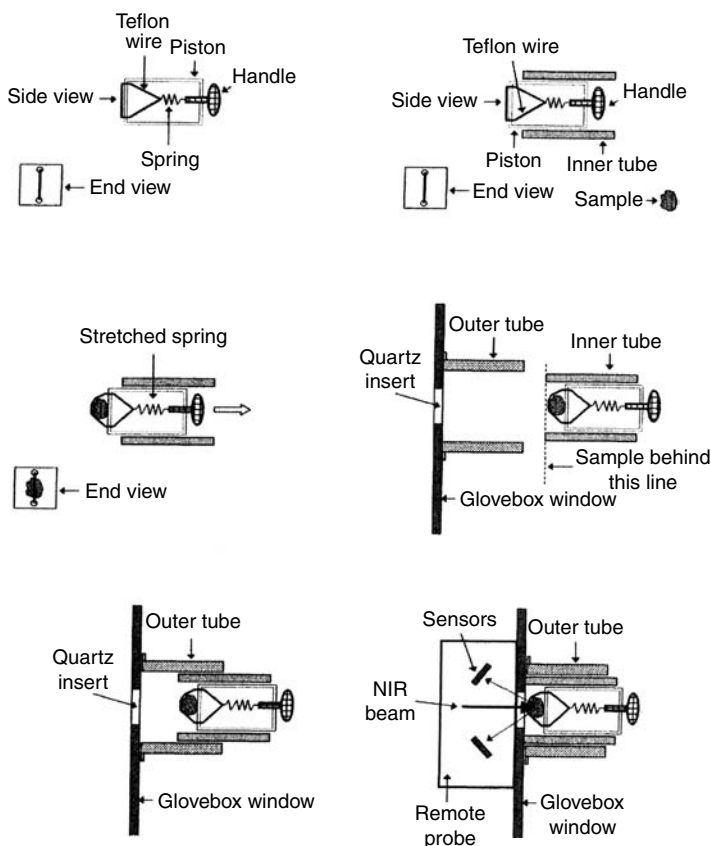


FIGURE 36.10 Special sample holder.

end of the piston by the Teflon wire (center left). The piston was pulled back until the sample could no longer be seen when looking across the end of the inner tube (center right). This entire assembly was then inserted into the outer tube, which was attached to the inside of the glovebox window and was also lined with black felt on its inner surfaces (bottom left). When the inner tube was inserted as far as it would go into the outer tube, the sample was positioned close to, but not touching, the quartz insert in the glovebox window. In this position, the NIR beam impinged upon the sample and

the reflected light was detected by the sensors in the remote reflectance probe attached to the outside of the glovebox window (bottom right).

A library of spectra was created by scanning 148 individual samples representing 16 categories of anticipated plastics and rubber. Scans of the following materials were taken over the range 1100 to 2500 nm:

Plastics		Rubber
HDPE [2]	PP [5]	Butyl
LDPE [4]	PS [6]	Latex
Nylon	Teflon	Neoprene
PET [1]	Tygon	Nitrile
Polyurethane	Polycarbonate	Vinyl
PVC [3]		

The numbers in brackets following some of the plastics are manufacturers' labels that are printed inside triangles molded into those plastics that are recyclable.

The push-button switch on the remote reflectance probe was rewired into a foot-switch so that the analyst, whose gloved hands were inside the glovebox manipulating each sample, could initiate a scan most efficiently. Each scan was displayed in real time on the computer's monitor in $\log(1/R)$ mode (absorbance vs. wavelength). Some samples were scanned more than once, and absorbances were saved at 2 nm increments (700 data points) in the computer and on a floppy disk. About 20 plastic or rubber samples were analyzed each day. If this number seems low, know that for each sample other people were making written records, audio recordings, and multicamera videotape records for complete documentation of the procedure.

At the end of each day, all the scans were subjected to qualitative analysis using the vendor's software. After evaluating various data treatments, first derivatives were used for all library entries and all predictions. The resulting printed report contained the date, time of each scan, sample name, sample identification when compared with the stored library, and Mahalanobis distance from the library sample. [Editor's note: if the reader doesn't understand Mahalanobis distances, refer to Chapter 15 in Part III.]

36B.3 RESULTS

A portion of a typical printout is given in the following table:

Plastics and Rubber Characterization STTP at Los Alamos National Lab			
Date of analysis		3-27-95	
Data points/scan		700	
Scan range		1100 to 2500 nm	
Method		Distance	
Time	Sample	Identified as	Distance
16:41	PVC?	PVC	2.3
16:48	?	Nylon	1.2
16:52	Nylon	Nylon	1.6
17:00	Syringe	Poly Prop	1.7
17:02	Thin sheet	PVC	2.1
17:08	Small bag	LDPE	2.0
17:51	Tape, msk	Latex	16.1

The entry under “Sample” is what the operator predicted when the scan was initiated. The output under “Identified as” is what the system software found after comparing the scan with those in the library. “Distance” is the Mahalanobis distance: if <3 , the identification was considered correct. For some samples, the distance was >6 (i.e., ± 6 standard deviations from the center of the cluster in the library) and the identification was considered incorrect. During the early days of the project, results such as that obtained at the 17:51 scan (when the sample was thought to be masking tape) prompted the addition of one or more standards into the library until the incorrect answer fell into a known category. Latex had not been included among the standards when the 17:51 sample was scanned. As the library expanded and the study progressed, the operator’s predictions improved.

Sometimes a library can be made more robust by adding to it the spectra of “unknown” samples that become known by reference to the literature. A case in point within this study is illustrated in Figure 36.11. *Pylox* gloves, according to the vendor, are made of vinyl rubber, but the analysis with the original library declared them to be unknown. When the two groups of spectra were compared, however, it became apparent that the only real difference was the water associated with the gloves (note the water peak at 1940 nm in Figure 36.11a [which is absent in Figure 36.11b]). So, when the spectra of the *Pylox* gloves were added to the library, the “fingerprint” for vinyl was modified (in the

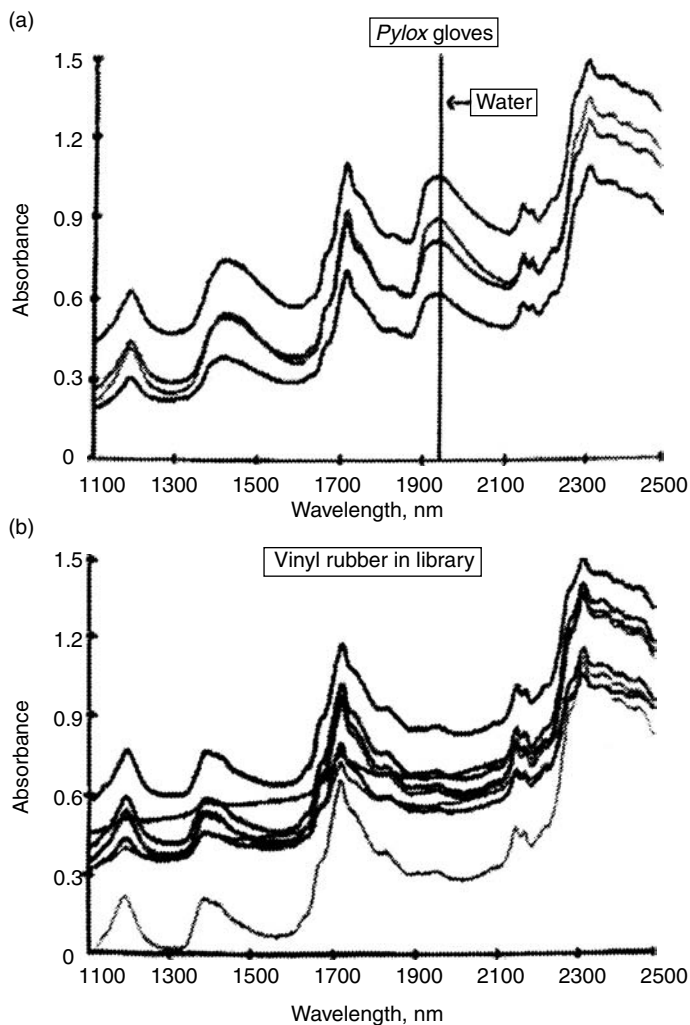


FIGURE 36.11 Effect of adsorbed moisture on spectra of *Pylox* gloves.

region of this water peak) to include a larger standard deviation to accommodate this added water, thereby enabling the software to properly identify the gloves as vinyl. Clearly, unknowns cannot be indiscriminately added to a library just so their identification matches the operator's wishes. There must be evidence that the unknown is in fact the same as that which was previously in the library.

The following tables shows the ranking of the materials analyzed by NIR. Note that this does not reflect the ranking of the total content of all 15 drums. Teflon appeared more often than anything else, and was eventually recognized without recourse to an actual analysis. The category "unknown" contains those materials whose spectra could not be found in the library when the study was terminated. This number was higher before some members of the group were placed in a category following operator (manual) scrutiny of their spectra:

Rank	Material	Percent
1	Teflon	19.1
2	Unknown	17.5
3	LDPE	16.4
4	PP	15.8
5	PVC	11.5
6	PS	5.5
7	Nylon	3.3
8	Latex	3.3
9	PET	2.2
10	Vinyl	2.2
11	HDPE	1.6
12	Polycarbonate	0.5
13	Tygon	0.5
14	Butyl	0.5

36B.4 DISCUSSION

It is typical of most NIR spectra to tilt upward at the high wavelength end of the spectrum. This baseline "tilt" can be removed by converting the spectra to one of their derivatives. For example, the family of teflon spectra in Figure 36.12a shows an average increase of about half an absorbance unit over the standard wavelength range 1100 to 2500 nm. This tilt is no longer observed in the first derivative display in Figure 36.12b. Not surprisingly, it is customary when building a library to make this kind of conversion.

In addition to flattening the baseline, this technique also removes much of the offset. Of course, this means that each spectrum of an unknown must be converted to the same derivative before a search of the library is made. Second derivatives are frequently used and sometimes are a route to better matches. In this particular circumstance, first derivatives are necessary and sufficient.

Another example, this time using a family of spectra with much more structure than that exhibited by teflon, is butyl rubber. Here, the initial absorbance is quite high (Figure 36.12c). This fact is "lost" when any derivative is taken (Figure 36.12d). Should the derivatives of two different materials be so similar that confusion could arise and no distinction could be made, then the operator might need to look at the original (underivatized) spectra to assign an identity.

36B.5 CONCLUSION

NIRS was applied in a nondestructive, noninvasive manner to characterize plastics and rubber materials in a large glovebox containing radioactive waste. A modification of the glovebox window, coupled

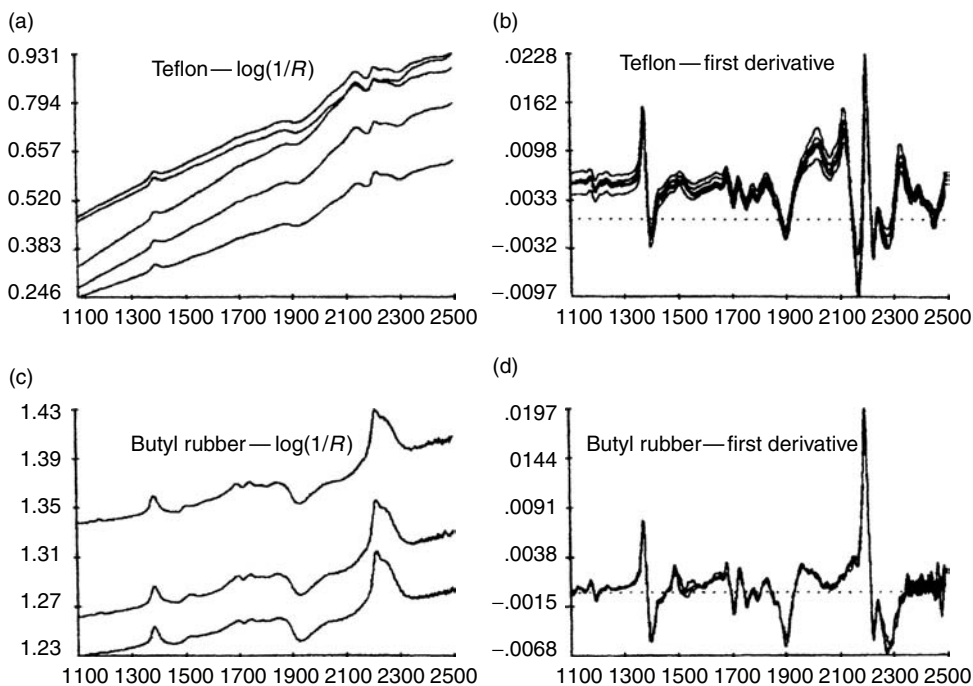


FIGURE 36.12 Effect of derivatizing the $\log(1/R)$ spectra.

with a special sample holder, precluded any chance of contaminating the commercial instrument. One hundred eighty-three samples were qualitatively analyzed over a period of one month, compared with a library containing 16 categories, and 82.5% were identified.

REFERENCES

1. D. A. Burns and R. Villarreal, Identification of Plastics in a Hot Glovebox, *35th ORNL/DOE Conference on Analytical Chemistry in Energy Technology*, Oct. 13, 1994, Gatlinburg, TN.
2. STTP Procedure for Characterizing Plastics and Rubber at WCRFF, Los Alamos National Laboratory, Publication *CST-STP-OPS5-061/0*, February 1995, Los Alamos, NM.

37 Process Analysis

Gabor J. Kemeny

CONTENTS

37.1	Introduction	717
37.2	Requirements of Process Analysis	719
37.2.1	The Analyzer and Its Environment	719
37.2.2	Rate of the Analysis	720
37.2.3	Analyzers in the Process Control System	721
37.3	Analysis of Liquids	722
37.3.1	Optimal Pathlength	723
37.3.2	Factors Influencing Liquid Analysis	725
37.3.2.1	Sample Transport Systems	725
37.3.2.2	Pathlength Variation	727
37.3.2.3	Effect of Temperature Variations	729
37.3.2.4	Filtering	731
37.3.3	Process Liquid Cells	734
37.3.4	Fiber-Optic-Based Analyzers	740
37.4	Contact Analysis of Solids	745
37.4.1	Sampling Parameters	745
37.4.2	Solid (Powder) Analyzers	747
37.5	Noncontact Analysis	750
37.5.1	Factors Influencing Noncontact Analysis	750
37.5.2	Sample Speed Variation	752
37.5.3	Noncontact NIR Analyzers	755
	References	757

37.1 INTRODUCTION

In the last 50 years, a new principle regarding competitive advantage has emerged, and it is having a profound impact on the way we think of products and production. In the center of this line of thought are the quality of the product, the quality of the production process, and the relationship between the two. According to the “0th law of quality management,” the producer of highest quality is also likely to be the producer of the least cost product [1]. This principle runs in contrast to the formerly generally accepted concepts in economy.

The role of quality is increasingly recognized in continuous chemical and food production as well as in pharmaceutical batch production and other processes. In continuous processes, from an analytical chemical approach, the goal in most cases is to keep the process composition steady at around the optimum physical and chemical conditions. Uniform quality is a requirement from many other aspects too. There are regulatory obligations for product composition, and in many cases, the most economical product is the one closest to the regulatory or label limits. There are requirements for environmental protection for production and plant safety. All of these require that the composition of

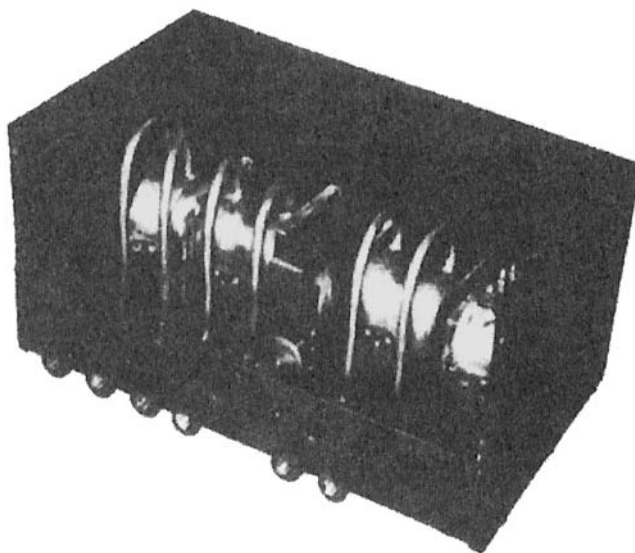


FIGURE 37.1 Infrared plant analyzer from 1943. (Courtesy of *J. Appl. Spectrosc.*)

various products remain stable. This principle was analyzed quite some time ago [2], and the theory of control, the controllers, and the actuators were developed and became available in parallel with the growth of the chemical-petrochemical industries. It was realized early, on that the “missing link” was the reliable, selective, and sensitive process analyzer.

Near-infrared (NIR) is one of the techniques that can be used to nondestructively analyze a product online at the process level. Because of the advantages of optical analysis, the development of infrared (IR) process instrumentation goes back to the early 1940s. Figure 37.1 shows a continuous infrared (CIR) plant stream analyzer from 1943; this was already a second-generation analyzer with improved electronics. Since then, the online optical analyzer instrumentation has grown into an established business sector within process instrumentation. The share of complex multicomponent analytical instruments started increasing (IR spectrophotometers, mass spectrometers, nuclear magnetic resonance (NMR) spectrometers, and different chromatographs) since the late 70s [3,4]. Various NIR laboratory techniques are also increasingly implemented in online food, chemical, petrochemical, textile, paper, and pharmaceutical applications. The examples of commercially available NIR process analyzers are included to illustrate the principles described. There are many other manufacturers and instruments with equivalent performance; examples and lists included in this chapter do not represent endorsement or selection by the author.

In most areas of production, the output of the plant must be monitored. The simplest monitoring occurs when random or systematic samples are pulled and taken “offline” to get the analysis results in a quality control (QC) laboratory. These results are then fed into a statistical process control algorithm or software in order to check the consistency of the plant output.

The next level in the control of the production output is a continuous auditing function done by an automatic analyzer. A continuous or semicontinuous auditing of concentrations of the finished products can alleviate the product liability problems, but cannot prevent the large-scale massive production of off-specification finished goods. This is why, in nearly all chemical manufacturing processes, there is a desire to analyze and monitor the intermediate products, not only the finished goods, and make process control decisions on the basis of the results of the analysis. This is in short the “closing of the loop” on the analyzer [5]. The same thinking, that is, recognizing the necessity to control processes by understanding and monitoring intermediate products, led to the process analytical technology (PAT) initiative of the FDA. In their guidance letter [6], the FDA stated, “The agency considers PAT to be a system for designing, analyzing, and controlling manufacturing

through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality.”

There are many publications discussing the needs for online analysis. Compositional analysis in a process is required to check the quality of raw or starting materials. It is also needed for closed-loop control: (a) to increase the efficiency of the processes, (b) to decrease the manpower requirement, (c) to ensure the safety and integrity of chemical processes, (d) to ensure the level of industrial hygiene, and (e) to monitor the by-product and waste outputs for environmental control.

Beyond the off-line analytical instrumentation — which is usually located in the plant QC laboratory — Callis and his coauthors [7] defined different eras of process analytical chemistry.

The first step is to take the instrument physically close to the process stream. This is called by Callis the *at-line* era, in which approach the analyzer is dedicated to the process line. Because it is physically closer and usually operated by plant personnel, the delay time of the analysis associated with the off-line instruments is drastically reduced.

The second is the *online* era, in which the instrument is connected to the process line and either is drawing intermittent samples, or is continuously sampling the process. NIR instruments have been extensively used in the QC laboratory, as most of the application chapters of this book would testify. The unmodified instruments are frequently used very close to the processes. Liquid samples are, for example, drawn from the process streams and are pumped through modified laboratory liquid cells.

According to the definitions given in Reference 7, *in-line* (the third era) usually means chemical or electrochemical interaction of the sample and the sensor that is in direct contact with the process stream, typically with the main process stream. This is an extremely advantageous approach for “difficult” samples, because the sample stream does not have to be diverted, filtered, or treated outside the process line. In fast, sensitive processes, the additional advantage is that there is no delay typically associated with bypass sampling. NIR analyzers can be configured to fulfill the definitions of the in-line arrangement. The fourth, the *noninvasive* era, means that the optical analysis done by NIR light does not destroy the sample and does not interact with the sample to change its composition; thus in both the liquid and the “noncontact” sample arrangements (see following chapters), the noninvasiveness criterion is fulfilled. In solid “contact-type” sampling, a grinding step could be introduced that changes the physical properties of the sample; this arrangement is therefore invasive to the process. The ground portion usually has to be directed to waste after the analysis.

Although NIR process analyzers can be classified according to the above “eras,” it seemed preferable here to classify them according to their sample presentation techniques.

The general requirements of process analysis will be discussed followed by the requirements and actual examples of liquid analysis, contact-type solid analysis, and noncontact sample arrangement. The spectroscopic and mathematical principles of NIR online analysis are similar to those discussed in the other chapters of this book. The emphasis of this discussion will be on the specific instrumental requirements and special sampling requirements that are unique to the process analytical approach.

37.2 REQUIREMENTS OF PROCESS ANALYSIS

37.2.1 THE ANALYZER AND ITS ENVIRONMENT

Process analyzers are integral parts of the process control loop. Therefore, their parameters have to match the requirements of the control loop of the particular process.

Large-volume continuous production processes are designed for 24-h operation. Therefore, the analyzers should also be capable of continuous uninterrupted operation. Well-designed analyzers are reliable enough to provide analytical results over long periods of time. In case they should fail, the built-in diagnostic routines detect the faulty conditions before the false result could damage the product through the misguided control system. As soon as the false operation condition is recognized, an error message is generated. This could be a digital message or a simple digital logic signal connected to the process controller receiving the analyzer concentration output. Fault conditions

occur, not only due to hardware failure, but also due to a faulty sampling condition. The latter is more frequent, and could be caused by clogging or bubbles in a liquid cell or by the temporary absence of sample material in the case of solid sampling.

The 24-h-a-day unattended operation is made possible by automatic sampling systems and intelligent supervisory diagnostics. The only planned interruptions in the operation of the analyzer should be the calibration/recalibration and the maintenance.

High on-stream availability requires that both the analyzers be protected from the plant/process environment and that the process be protected from the hazards the analyzer may generate.

NIR analyzers do not use flammable diluents, reagents, or carrier gases, but the light source, which in most instruments is an incandescent tungsten–halogen lamp operated at high temperature. The lead-salt photoconductive NIR detectors require up to 200 VDC, which represents electrical spark hazard if the instrument enclosure is not designed appropriately.

In general, however, optical analyzers pose less danger to the process environment than the plant environment does to the analyzers. Acidic fumes cause irreversible damage to various optical elements — mirrors, filters, gratings, and beam splitters. Sensitive mechanisms, such as grating drives, filter wheels, and interferometers, are also in jeopardy of corrosion. In NIR analysis, the analytical signal change of weak absorbers is often below one milli-absorbance. The slightest deposition of foreign materials, delamination, and pitting of the optical elements could cause a change in the detected signal and result in gradual change of calibration.

The internal optical/mechanical elements can be protected by proper mechanical and systems design. The enclosure designs have sound well-tested practices. To support this, there are design and testing guidelines, standards, testing companies, agencies, and consultants with specialized expertise in this area.

Some effects of the environment or the sample itself cause optical changes exterior to the instrument enclosure. Sample can deposit onto the optical window of the sample cell, or water can condense onto the optical window of a noncontact analyzer from vapor phase. Remedies for some of these types of problems will be discussed later under the respective sample types.

37.2.2 RATE OF THE ANALYSIS

Optical analysis can be done relatively fast. NIR analysis can be performed within a few seconds, depending on the magnitude of the analytical signal, the sample absorbance, and the overall error of the measurement. Simple analytes like moisture require only two or three wavelengths. On the other hand, more complex analytes may be best calibrated using up to six wavelengths. Not only does the scanning at more wavelengths take more time, but the smaller analytical signal associated with the weaker absorption also requires longer averaging.

The purpose of the analysis is to use the data to control the process. The processes have characteristic transfer functions that govern the rate of change due to compositional alterations or due to change in the control settings. Speed of analysis is usually not a limiting factor in slowly changing production of bulk products. A large drying kiln or a reactor vessel has a large time constant and lag time within which the control change could be detected in the output. Some high-speed film or web defect detection, however, requires faster optical arrangements and special high-speed analyzers. Defects in plastic films and fibers could show up within milliseconds. The analyzer must be able to monitor the change in composition as it occurs, without delay and without distortion. If fluctuation occurs with a frequency f , then according to the Nyquist criteria, the sampling rate (rate of analysis) has to be at least $2f$ to be able to monitor the change. In practice, an approximately $4f$ to $6f$ frequency is recommended for an undistorted characterization of the process variable. Honigs [8] discussed the timelines of NIR analysis.

Figure 37.2 illustrates the relationship of a changing process signal and the frequency of the analysis. If the large swinging of the signal (a) is to be corrected, the analyzer has to be able to detect the changes on the same time scale as it occurs.

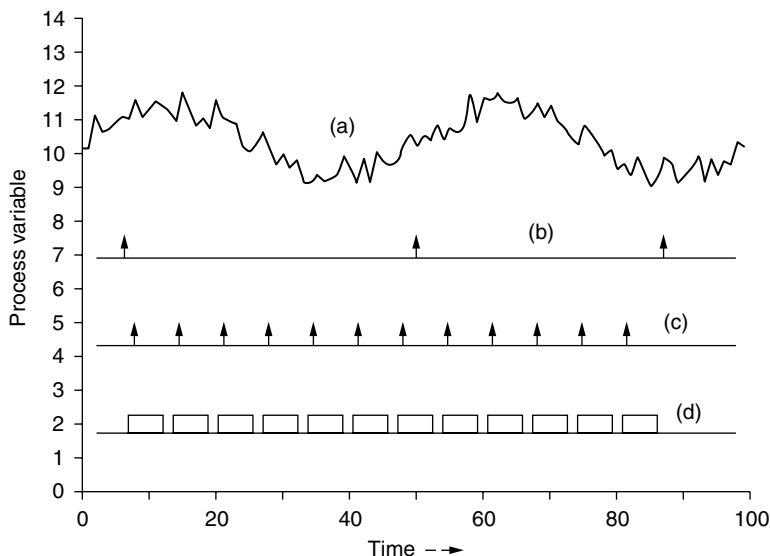


FIGURE 37.2 Relationship of process variable and the sampling rate: (a) process variable (concentration), (b) under-sampling, (c) appropriate sampling rate, and (d) sampling with signal averaging.

Analyzing the process variable signal, a large sinusoidal variation is observed. This remains unnoticed with the point sampling of scheme (b). Sampling scheme (c) provides enough measurement points to follow the signal's low-frequency variation. There should be no control action taken at each random local variation of the signal because the point sampling also carries the errors of the local variations. The ideal analyzer, shown as sampling (d), covers the whole available time, averaging the individual readings to decrease the errors of the local variations of the process variable, which is, in this example, the concentration. Averaging a plurality of readings at the same time decreases the sample representation error and the instrumental error of the analysis simultaneously.

The string of analysis output results mathematically in a time series data, usually with even time spacing. The mathematical treatments developed for time series analysis can be applied to predict trends, drifts, etc., which have impact on process control decisions. The analysis results can form the input data for statistical process control, developed earlier for discrete production [9,10]. Figure 37.3 shows one of the basic types of presentation charts. Statistical process control has an extensive literature (see Reference 11).

37.2.3 ANALYZERS IN THE PROCESS CONTROL SYSTEM

The process analyzers are not only physically and environmentally integral parts of the processes, but they have to be part of the same information-processing system. The block diagram of Figure 37.4 indicates the central role of the result that the analyzer provides. The process controller (PLC), the analyzer, and the actuators (not shown) form the local loop, which has to work uninterrupted in a continuous production.

Local and remote concentration displays provide information to the operators of the process, which allows an off-line control loop based on human decisions.

Remote central computers supervise the whole production process, observe and register the trends, note outlier data points, compute statistics, and so forth. One of the supervisory computer functions is the monitoring of the performance of the analyzers in order to organize the timing of the calibration sessions.

There are several other factors that might influence the operation of the analyzers that will not be dealt with here (vibration, EMI/RFI shielding, etc.). These and other details of the features,

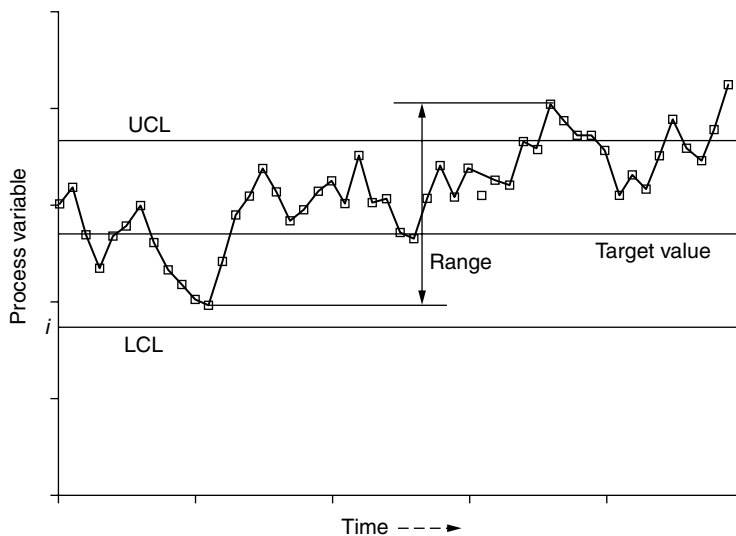


FIGURE 37.3 Presentation of process variable.

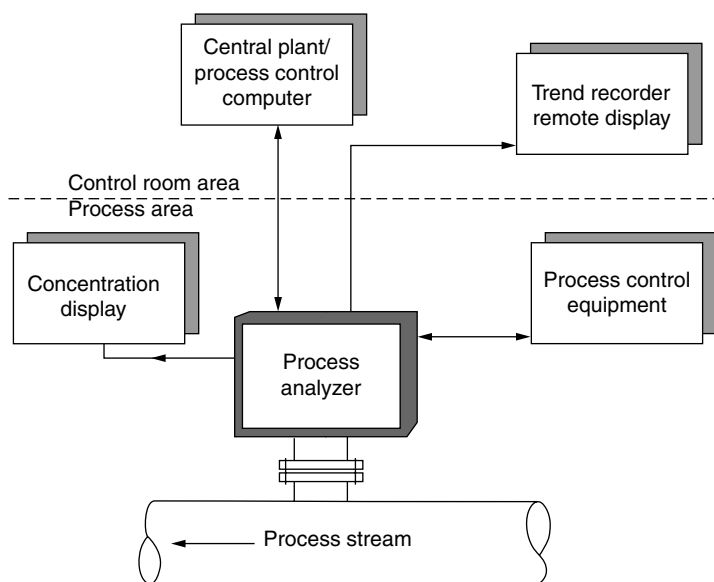


FIGURE 37.4 Process analyzer in process control.

requirements both technical and organizational, are compiled in many articles and books [12–15]. Some of the first NIR online installations were reported in the early 80s [16,17].

37.3 ANALYSIS OF LIQUIDS

NIR analysis of liquids has the longest history of all sampling types; simple two-filter, two-wavelength instruments have been used for decades, mainly for the analysis of moisture in various liquid media. The sample presentation of liquids seems to be much simpler than that of solids, but there are several factors influencing the accuracy and precision of the analysis that need to be addressed for a successful liquid analysis.

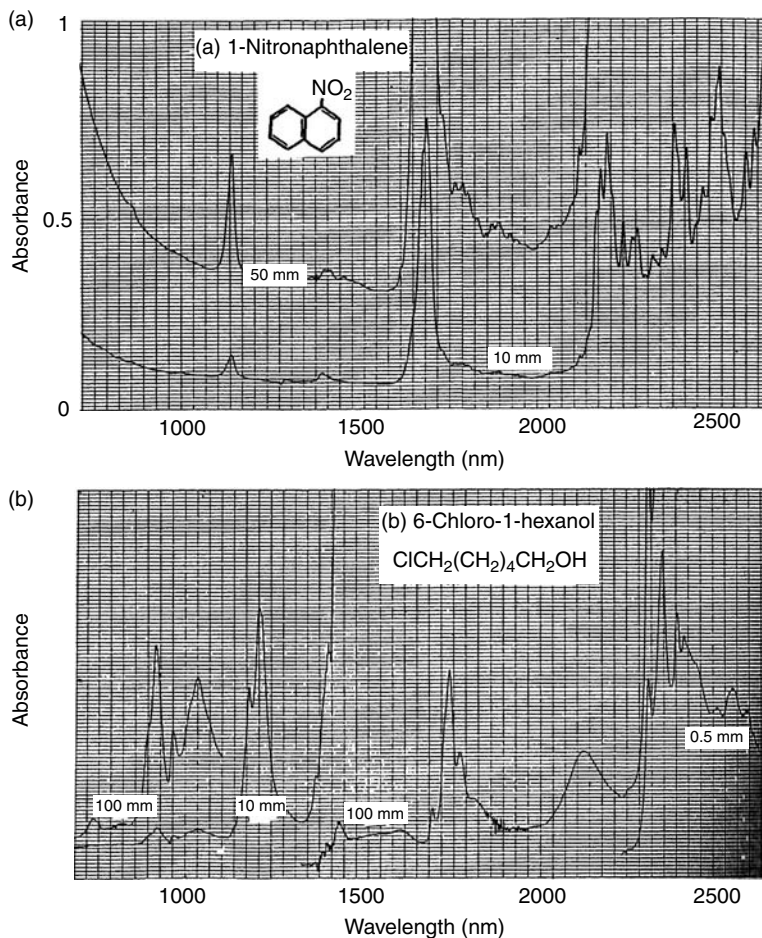


FIGURE 37.5 Transmission spectra of liquids: (a) 1-nitronaphthalene and (b) 6-chloro-1-hexanol. (Courtesy of Sadtler Research Laboratories.)

37.3.1 OPTIMAL PATHLENGTH

In Figure 37.5a and Figure 37.5b, the spectra of two arbitrarily chosen liquids are shown in the 700- to 2600-nm region. The —OH-containing material has much higher absorbance and broader peaks in the whole spectrum. In both spectra, similar to other substances, the higher order overtones and combination bands are weaker following an approximately logarithmic scale. In the mid-IR region, it is typical to record the IR spectrum of an organic material in a film pressed between two mid-IR transmitting disks that limit the thickness of the film to a few micrometers. However, in the long-wavelength part of the NIR spectrum, typical liquid cell pathlengths can be as large as a few millimeters. On the short-wavelength side of the spectrum though, the absorption bands are so weak that in order to achieve a few tenths of an absorbance unit, the cell pathlength has to be increased from 10 to 100 mm.

One of the important parameters of liquid analysis is the optimum selection of the pathlength to achieve the most precise analytical results. Examining Figure 37.5, it is obvious that different optimum pathlengths apply, depending on the wavelength range. In classical spectrophotometry [18], it was known that the optimum sample absorbance is around 0.434. It is less known that this value is optimum only under several assumptions. One of the assumptions is that the spectrophotometer noise is detector noise-limited. In other words, all other factors influencing the noise, and thus

the measurement, are significantly smaller than the absorbance or transmittance imprecision, due to detector signal fluctuations. The other assumption is that the material is analyzed by a single-wavelength determination.

NIR analysis is typically a secondary analytical method, that is, it has to be calibrated with several samples of known concentrations. The optical readings at several wavelengths are combined with the known concentrations to provide multiple-wavelength absorbance readings. The optimum pathlength in a multiwavelength analysis was discussed by Honigs et al. [19]. In a linear multiwavelength calibration the concentration is determined by

$$C(\%) = K_0 + K_1A_1 + \cdots + K_nA_n \quad (37.1)$$

where K_0, \dots, K_n are calibration constants, and A_1, \dots, A_n are measured absorbances at different wavelengths. Expressing the absorbances in transmittance, we obtain:

$$C = K_0 + \sum_{i=1}^n K_i(-\log T_i) \quad (37.2)$$

Differentiating equation (37.2) and expressing it in a natural logarithm:

$$dc = \frac{1}{2.303} \sqrt{\sum \left(\frac{K_i}{T_i} dT_i \right)^2} \quad (37.3)$$

is obtained. The error of the concentration reading is added from the individual error terms at each wavelength. There are no cross-error terms because the optical measurements at each wavelength can be considered statistically independent. The goal of this analysis is to minimize the relative concentration error. Assuming that the detector noise E is constant throughout the analysis wavelength range and it is the dominant error source, all transmittance errors (dT_i) are assumed to be the same and equal to E . Rewriting the formula in absorbance:

$$\frac{\Delta c}{c} = \frac{E}{2.303} \sqrt{\frac{\sum K_i^2 \times 10^{2A_i}}{\sum K_i A_i}} \quad (37.4)$$

Some simplified cases can be considered. In the first case, all absorbances of the material are the same. This is a relatively good assumption, NIR spectra being relatively featureless. The minimum concentration errors occur at the classical optimum absorbance of 0.434 in this boundary case. In another special case, one term is considered much larger than the other ones, this being the case of one strong absorbance band dominating the spectrum with the other wavelengths used as references. Equation (37.4) is reduced to

$$\frac{\Delta c}{c} = \frac{E}{2.303} \frac{10^{A_i}}{A_i} \quad (37.5)$$

The optimum value again is 0.434.

When the entire spectrum is used for the calibration, the optimum pathlength is obtained if the mean absorption approximates 0.434 rather than any individual absorbance reading.

The last simplified case is that in which several wavelengths are used in the calibration. Optimization calls for a pathlength where the minimum absorbance results in error equal to the error introduced by the maximum absorbance included in the calibration equation. Based on the above, NIR analysis has the same pathlength optimum as other optical methods. NIR detectors have very good noise performance, thus it must be emphasized again that these conclusions are only valid if the system is detector noise-limited and not amplification-, analog-to-digital conversion-, or pathlength-limited, which may be the case in some NIR process applications.

37.3.2 FACTORS INFLUENCING LIQUID ANALYSIS

37.3.2.1 Sample Transport Systems

On-stream process analyzers are known to be no better than their sampling systems. This emphasizes the importance of a well-designed sampling arrangement, because in the practice of the process analysis most of the downtime and false readings occur due to sampling failures rather than electronic or instrument malfunction [20,21]. A reliable sampling system is a must in NIR analysis, too.

The liquid samples can be actively or passively transported through the liquid cell. Active transport in this sense means the addition of a pump in a bypass line, whereas passive transport is driven by pressure difference. The pressure drop across the sample cell could range from a few mm Hg to several bars. If the pressure drop required for the sample flow can be maintained by the pressure difference between two sampling points in the process, the sample will flow through the cell without an external pump. This could be achieved by inserting a restriction in the flow of the process material if the overall pressure requirements allow this pressure drop. The two sides of the orifice will provide the necessary pressure difference as long as the material is moving in the main line. This is a relatively simple way to achieve the sample bypass loop, but the stream in the main pipe must be constant, otherwise the pressure difference will vary with the rate of material flux in the main pipe. Some basic sample transport arrangements are shown in Figure 37.6.

In some processes, where the reaction kinetics or the possible contamination of the sample does not allow the material to be fed back into the process stream, the sample is directed to waste, and consequently the outlet pressure drops to atmospheric. This may allow a simple, single-line sampling transport from the process if the pressure at that point is above atmospheric. The drawback of this type of sampling is that an aliquot of the process material is wasted, and if it poses environmental hazards, it must be properly disposed of or recovered.

The global pressure in the process could be low, close to atmospheric, for example, in an open reaction vessel. In this case, an external small pump (e.g., a peristaltic pump dedicated to the analyzer) could be used (Figure 37.6a). If the material is not homogeneous in the reactor vessel, the sample drawn from the bottom may not be the best option. Flange fitted sampling from the top of the reactor vessel, which could be adjusted to the optimum immersion depth, is possible. An external pump circulating the fluid drives the sample through the analyzer cell, re-circulating the fluid back to the reactor. Immersion depth in this sampling arrangement must be determined based on the maximum priming depth of the particular pump based on the viscosity of the fluid.

Material will flow through the cell if a pressure difference builds up due to the restriction in the main stream (Figure 37.6b). Transport can also be achieved by placing the sample inlet and outlet points on two sides of a process pump. The pressure difference is relatively constant while the pump is in operation (Figure 37.6c).

In principle, in all of these sampling arrangements, the sample could be continuously flowing or regularly stopped for each measurement. The continuous flow is simpler and requires less attention. In a stopped mode, the sample is more stable. However, particles could settle onto the optical windows during the static state. If the application calls for the stabilization of the temperature of the sample, the increased amount of material transported through the cell requires more heat transport to achieve the same controlled temperature range. A slow continuous material transport can be better stabilized for constant temperature.

All liquid cells using a small sample pathlength have very limited material capability throughout. This could cause significant lag time between the actual process concentration change and the analyzer detecting it. Reduction of the bypass volume is desirable to minimize the lag time. Reduction of the dead volume of the sample cell reduces the washout time and the back-mixing of the old sample with the fresh.

No external pipes or pumps are needed if the optical system allows in-stream operation. In this mode, the material is transported as it would be otherwise at that point of the process, and the

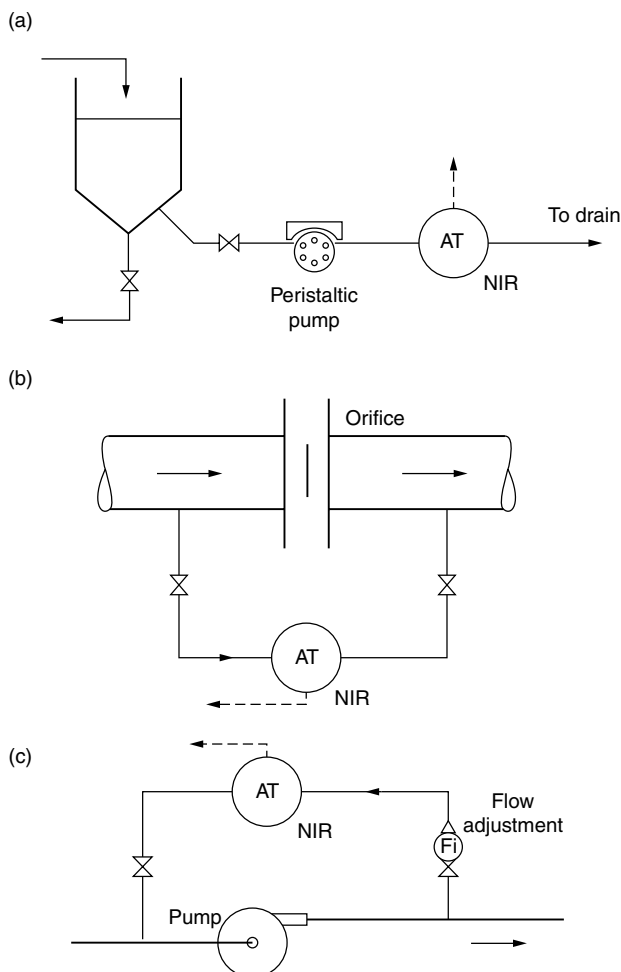


FIGURE 37.6 Sample transport systems: (a) in-line pump, (b) pressure difference (restriction of flow), and (c) pressure difference (pump).

analyzing light enters and exits through windows mounted in the pipe. Another in-line arrangement is an “optrode” protruding into the stream from one side of the process pipe or vessel. Most of these types of NIR analyzers are fiber optics based and will be addressed in the fiber optics section.

Figure 37.7 is a diagram of a typical bypass analyzer arrangement. The process line is sampled by opening the isolation valve. The flow can be maintained through continuous or intermittent flow sampler systems. Between the isolation valve and the sampling system, off-line samples can be drawn through a valve for laboratory calibration purposes. The sample conditioning system maintains that all physical parameters of the sample are appropriate and best for the analyzer. Sample conditioning and other important factors will be discussed in the following sections. The analyzer can operate between its specified pressure limits and yield the best results if the pressure is maintained relatively constant. A pressure relief and pressure protection line must bypass the analyzer. External calibration liquids may be introduced into the analyzer through a valve for *in situ* calibration. In many processes, even a small aliquot of the sample stream may be very expensive or may be very hazardous if directed to waste. Consequently, it is better if a sample recovery system retrieves and returns the sample to the process at an appropriate location.

An important factor of the sample transport system is the flow rate at the location of the optical analysis. The best results can be obtained with a laminar liquid flow. At higher Reynolds numbers,

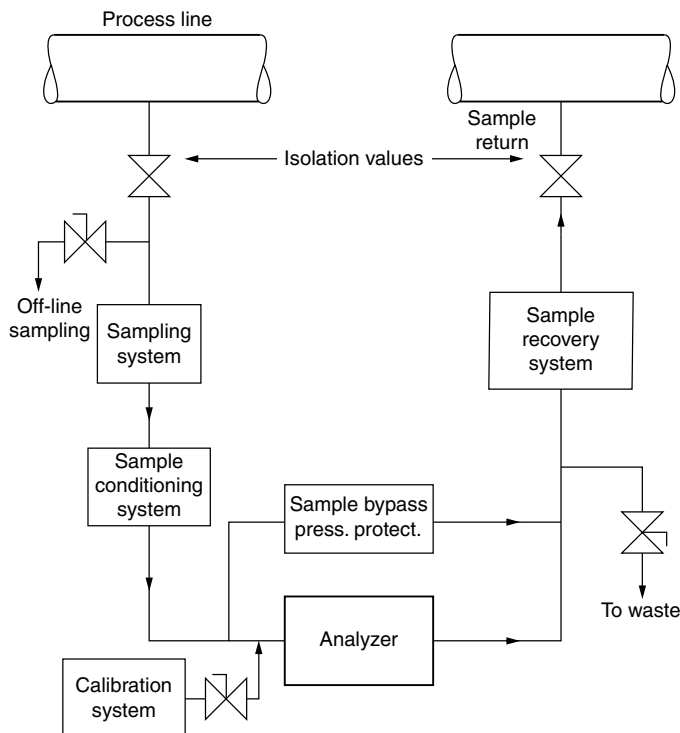


FIGURE 37.7 Bypass sampling system.

the liquid may form small currents that will result in noise added to the analytical signal. Even if the liquid is a homogenous matrix, local index of refraction changes could stem from small temperature variations, affecting the signal (Figure 37.8).

When dealing with multiple streams of process fluids, it is very important to keep the integrity of streams intact. In order to avoid the hazards involved with accidental mixing of streams, the single valve systems are not enough. A single-fault of a leaking valve could create a dangerous situation; this is why the so called “double-block-and-bleed” systems are used (Figure 37.9). This system consists of at least two valves in between any two pressurized points containing process liquids. Between the double valves is a bleed line with its additional valve, directing to waste the process fluid portion in the currently inactive stream.

In summary, the parameters of the process and the material determine which of the following is the desirable sample transport system:

- Active transport (pump)
- Passive transport (pressure difference)
- In-stream analysis (no bypass transport).

Each arrangement has advantages and disadvantages. The optimum sampling system, therefore, must be tailored to the actual application.

37.3.2.2 Pathlength Variation

One of the key parameters of a NIR liquid analyzer is the reproducible pathlength. This is pivotal in all optical instruments (spectrophotometers), but the relatively small pathlength — especially above about 1600 nm for more absorbing liquids (aqueous solutions) — renders this parameter critical.

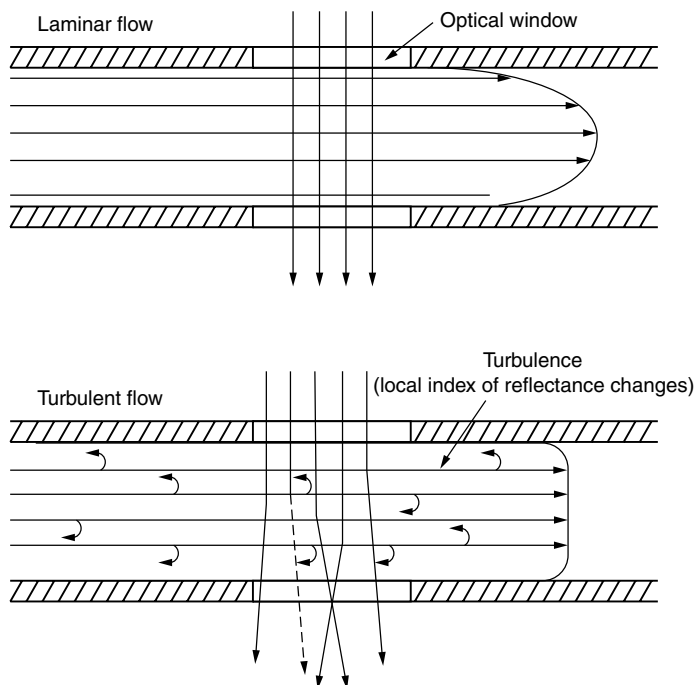


FIGURE 37.8 Laminar and turbulent flow in a liquid cell.

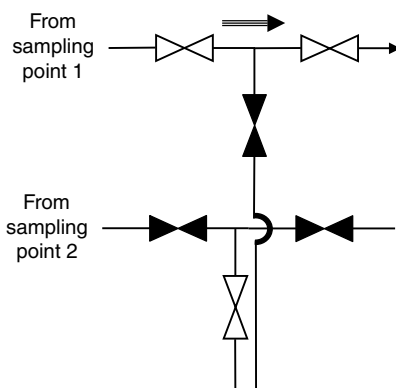


FIGURE 37.9 Double block and bleed concept to protect the process system from stream unwanted mixing due to single-fault failure (black = closed valve, white = open valve).

Furthermore, the pathlength of the cell should not change after dismantling the cell for cleaning and installing it again in the process. The pathlength should also be insensitive to slight pressure variations. At long wavelengths where the pathlength is very small, even the smallest dimensional change represents a large relative pathlength error. As discussed earlier, the optimum sample absorbance for quantitative analysis is dependent not only on the absolute absorbance values, but also on the other noise sources. The contribution of the pathlength error in photometric determination is treated in Reference 22. The concentration error is

$$\frac{\Delta c}{c} = \sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta b}{b}\right)^2} \quad (37.6)$$

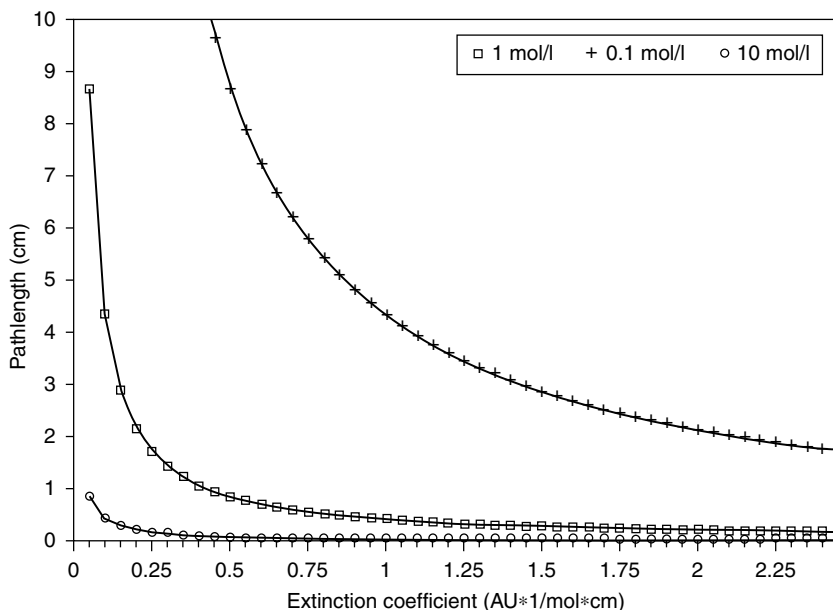


FIGURE 37.10 Optimum pathlength in ideal one-wavelength concentration measurement.

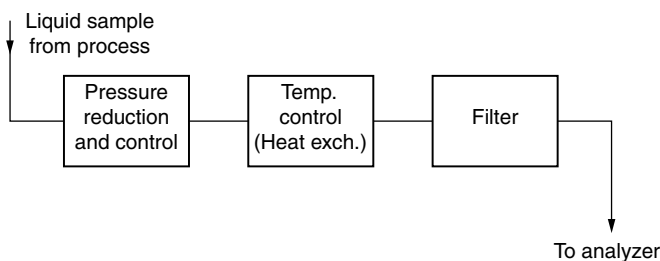


FIGURE 37.11 Sample pretreatment functions.

where $\Delta A/A$ is the relative absorbance error and $\Delta b/b$ is the relative pathlength error. Substituting the individual error terms does not lead to an analytically solvable equation. However, numerically calculating the relative concentration error as a function of absorbance, the optimum absorbance moves to a larger value. The optimum absorbance, as a function of the log of the pathlength error divided by the photometric error, is close to linear over three orders of magnitude, pushing the original optimum absorbance value from its original 0.434 to close to 2.0 AU. The $1/x$ type function for the simplest case with 0.434 as optimum absorbance is seen in Figure 37.10 as a function of molar concentration.

37.3.2.3 Effect of Temperature Variations

In most processes, the liquid sample has to be conditioned. In the following sections, the effect of temperature, pressure, phase, particle content of the sample, and the flow characteristics will be reviewed. Sample conditioning means that the sample fluid is brought to a condition that does not adversely affect the analyzer and that allows reproducible analysis results.

Conditioning, as shown in Figure 37.11, includes reduction and control of the pressure. This is required to protect the analyzer and to maintain a steady optical pathlength. Other conditioning modules are the temperature controller and the filter. It is known in the laboratory that for NIR

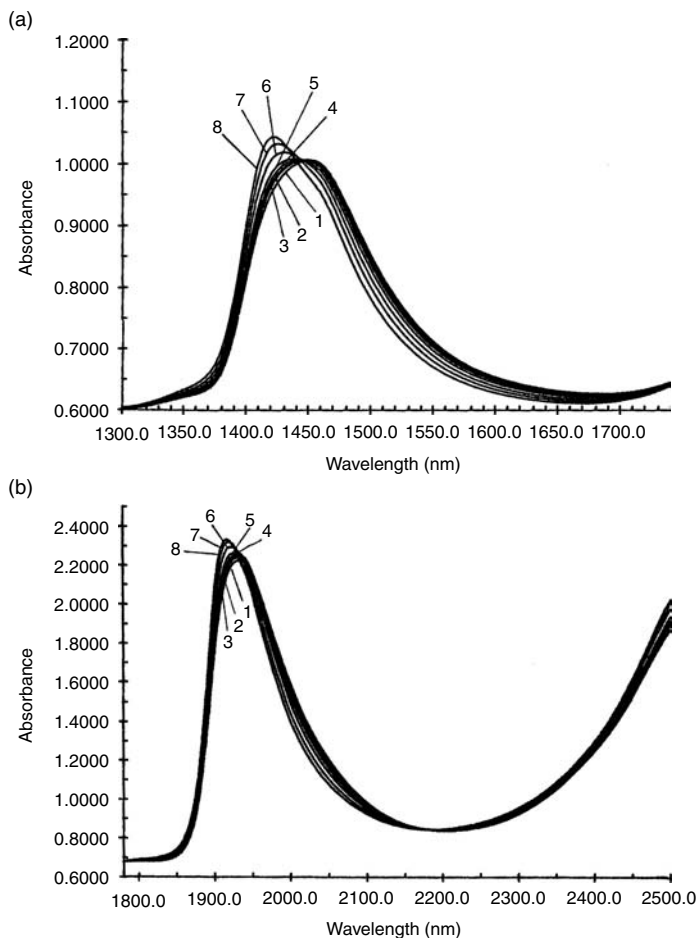


FIGURE 37.12 Spectra of water at different temperatures (°C): 1 = 10, 2 = 22, 3 = 28, 4 = 36, 5 = 44, 6 = 57, 7 = 74, and 8 = 86.

liquid analysis the samples must be maintained at a stable temperature in order to achieve precise concentration data. Slight temperature variations are known to change the hydrogen bonding structure of water in most cases, but the spectral characteristics of other materials through changing the hydrogen bonding or other chemical equilibria have also been observed. This is the reason laboratory NIR liquid analyzers have temperature-controlled liquid cells, ensuring approximately ± 0.1 to 0.2°C temperature stability. The spectrum of water (Figure 37.12a and Figure 37.12b), for example, is drastically changed by varying temperatures. In this example, the temperature was varied between 10 and 84°C . The set of curves shows strong changes in the spectrum, due to the appearance or disappearance of several water species. Liquid water in this temperature range has large molecular clusters — dozens of H_2O molecules held together by hydrogen bonds. The spectral changes are gradual, as the average size of these clusters becomes smaller with the elevation of the temperature. The intensity of some bands increases at the expense of others, resulting in shifts of the band maxima. Similar changes in the spectrum of water can be observed not only as a function of temperature, but as a function of absorption coverage affecting the same hydrogen-bonded cluster formations [23,24]. Needless to say, the strong effect of temperature change in aqueous samples can easily overwhelm the weak analytical signal in dilute solutions.

Within a range of a few degrees, where the effect of the temperature change can be considered linear, the calibration algorithms can correct these changes somewhat. The principle of using

temperature as an additional variable in the calibration is demonstrated in whole-grain analysis in a commercially available bench-top instrument (Trebtor Industries T-90). Outside a certain range, the sample temperature cannot vary without the potential loss of accuracy of the chemical analysis, even using temperature correction algorithms.

In laboratory-type instruments where the material flow is very small (e.g., a few ml/min) or where the flow is stopped before the optical measurement begins, there is ample time for the temperature to stabilize close to the targeted controlled temperature. In process analysis, an efficient heat exchange must be applied. In the in-line arrangement, where the main process stream is temperature-controlled, the actual temperature can be measured, and this value can be used to correct the calibration.

The other effect of temperature variation is that it changes the index of refraction of the liquids, which may change the effective pathlength of the liquid cell. Temperature control units are used in the sampling lines drawing a portion of the sample stream. Passive cooling can be achieved by an exposed section of the heat-conducting pipe; active cooling can be achieved with a small heat exchanger with cooling fluid. For stable sample conditions, the temperature should be maintained throughout the year, thus the exposed sampling lines may need to be heated with steam or an electrical heater in the winter.

Yet another aspect of the temperature in process analysis, is that some samples must exceed a certain temperature in order to flow (plastics, hydrogenated oils, etc.). In this case, the sample cell must be designed to avoid cold spots where the material could be deposited, eventually clogging the cell. In food processes, the deposition of the material is also highly undesirable for sanitary reasons.

37.3.2.4 Filtering

One of the limitations of NIR analysis of industrial samples is that the measurement is grossly affected by suspended particles and bubbles in the liquid. An aliquot of the main stream can be continuously filtered before it is introduced to the analyzer. This is especially required in the long-wavelength range of the NIR, where, due to the small pathlength, the liquid cell has to contain very narrow channels of a few tenths of a millimeter. Figure 37.13 shows an outline of a continuous bypass filter that can be used to treat the incoming sample stream and produce a small, filtered stream for the analyzer.

The homogeneously dispensed particles of a few micrometers diameter cannot be filtered out easily, but if their concentration and particle-size distribution remains relatively constant, the measurement can still be done, especially if the pathlength is small.

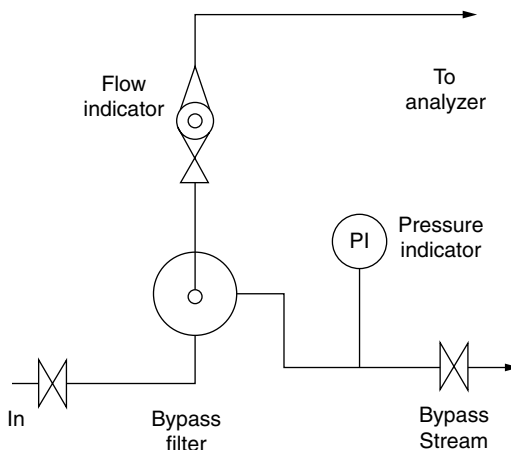


FIGURE 37.13 Bypass filter in analyzer stream.

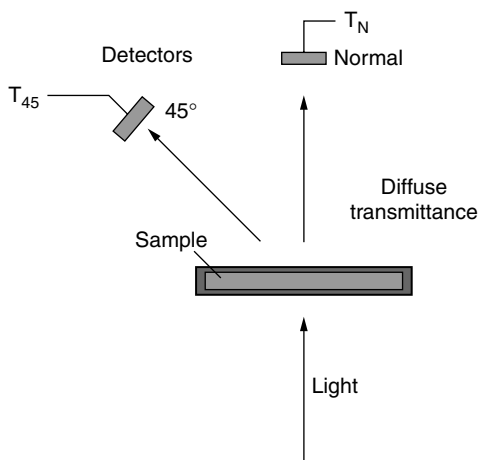


FIGURE 37.14 Schematic of apparatus to study light scattering in diffuse transmitting material.

The effect of suspended solids or other scattering material, such as fat globules, was studied by Birth [25]. Figure 37.14 shows the schematic diagram of his apparatus, where the light is impinging on a sample containing light-scattering centers. Two detectors are placed on the opposite side of the illumination: one detector on the optical centerline, the other at a 45° . The first detector measures the on-axis transmittance, which decreases as the scattering reduces the energy collected by the detector. The other detector measures the light intensity scattered by the particles in the 45° direction. The concentration of a dye in the presence of a scattering media, such as milk dissolved in water, shows nonlinear concentration dependence. This nonlinearity can be compensated using scatter-corrected absorbance (K'). The diffuse thickness δ can be determined with the simple apparatus, and can be used to correct absorption coefficients and calculate the corrected absorbance.

$$K' = \left(\frac{K}{S} \right) S' = \frac{K}{S} \times \frac{1}{\delta} \quad (37.7)$$

In a real sampling situation, however, different diffuse thickness values must be used for each individual sample set, and can only help linearize the concentration function if the degree of scattering remains constant.

Another disturbing factor in liquid sampling is the appearance of gas bubbles. It is known that liquids dissolve significant amounts of gases from the atmosphere. Even more can be dissolved if pressurized, as is usual in the production of beer or nonalcoholic beverages, for example. The gas is freed to form bubbles by the increase of temperature or the decrease of pressure. Both of these may occur in a real sampling situation when the equilibrium is changed. The pressure, for example, must be dropped for analysis if the process is operating at a higher pressure than that for which the liquid cell was designed. In some food processes (e.g., the production of beer) the material is cooled and bubbles will form in the sample if the temperature is elevated to ambient.

Some thin liquid cells were constructed to reduce the effect of bubbles by special hydraulic design. One such cell has a channel around the optical area to trap and direct the bubbles away from the optical beam (Figure 37.15).

Usually, the small bubbles are the problem because they may slightly change the optical reading. When a large bubble is trapped in the optical beam, it changes the light intensity enough to be a recognizable error condition. Random spikes of similar magnitude on the control chart can, in many cases, be traced back to bubbles trapped and later washed from the liquid cell. Small bubbles can be collapsed and separated from the liquid in bubble separators (gas separators) (Figure 37.16).

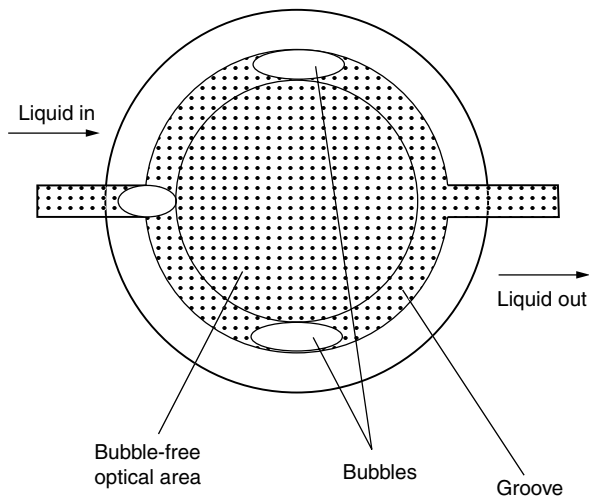


FIGURE 37.15 Bubble-free liquid cell.

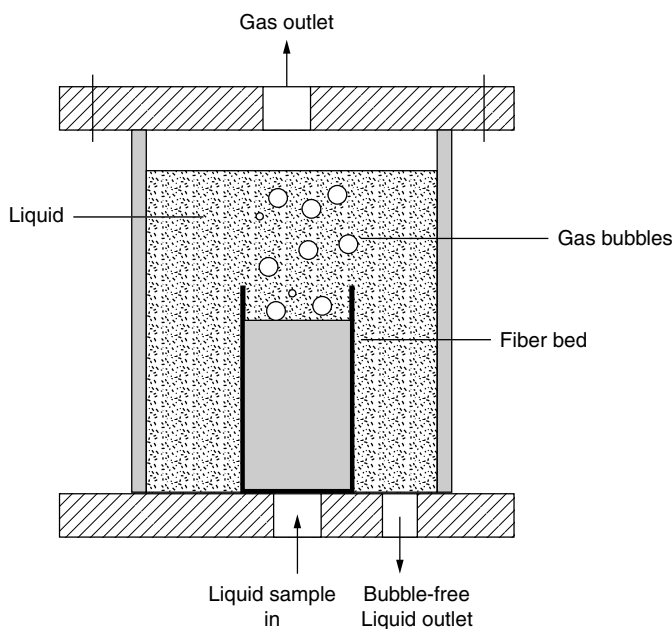


FIGURE 37.16 Bubble separator.

Yet another factor disturbing the optical liquid analysis could be an incomplete phase separation. Small globules of another liquid phase may cause subtle optical changes, less than the gas bubbles, because the index of refraction and the density of liquids are much closer than those of any gas. Phase separators of various sizes and efficiency are available and are typically well known in the particular industry, where the analyzer is to be installed.

In-line filters provide adequate solution only if the stream has a small number of particles (rust, catalysts, etc.). If the flow of material to be filtered is larger, or if it contains more suspended material, a bypass filter (Figure 37.17) may be needed.

Bypass filters of different designs are available. It is advantageous if the filter cleaning is continuous. Swirling of the liquid in the filter cavity is used in some commercial filters (Collins

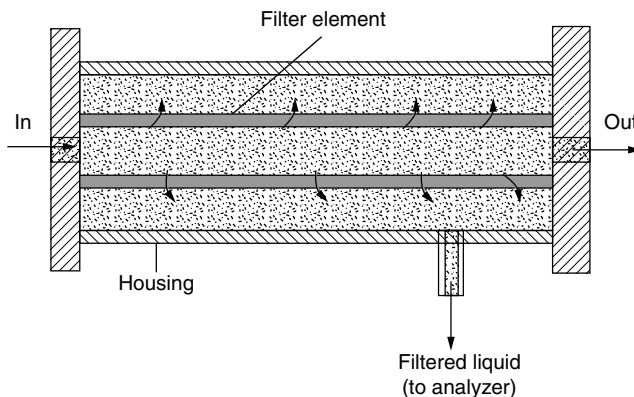


FIGURE 37.17 Bypass filter.

Products Co., Livingston, Texas). Different filters are available to filter out particles routinely down to $1\text{ }\mu\text{m}$ diameter. In other more complex filters, motor-driven agitation helps increase the efficiency of the filter. The centrifugal action not only helps to separate heavier particles, but also separates residues of the heavier phase. This latter feature is helpful in removing the water traces from gasoline-type streams.

The liquids in different processes are very different, not only in their pressure, temperature, viscosity, and flux, but also in their optical properties. As discussed earlier, the presentation of the liquid sample to the optical beam is the most important aspect of the successful analysis. Adequate sample conditioning systems will reduce the measurement error stemming from the temperature and pressure variations. In-line filters can reduce or remove the residual particles that would scatter the light passing through the liquid.

Sample conditioning is currently accomplished using general purpose discrete conditioning devices for all of the previously discussed issues of temperature, pressure, particle removal, phase separation, and flow [26,27]. Conditioning devices can be mounted on a common base plate and connected by standard connectors. This type of sample conditioning arrangement is rugged, and is usually dedicated to the application for which it was designed and developed (Figure 37.18 and Figure 37.19).

In order to achieve flexibility, faster development, and to reduce the design and installation cost, a new standard, the ISA/ANSI 76.00.02 was developed [28]. This system defines a modular sample conditioning system that can be assembled on a manifold plate using components that follow the size and connectivity protocols in the ISA/ANSI standard. Companies such as Parker Instrumentation (Jacksonville, AL) (Figure 37.20) and CIRCOR Instrumentation (Joliet, IL) (Figure 37.21) produce pressure reducers, gauges, flow meters, different types of valves, safety devices, and many other components that comply with the footprint and connectivity definitions of the ANSI/ISA 76 standard.

Another standardization proposal, the NeSSI (New Sensor-Sampling Initiative by CPAC, Seattle, WA) defines not only the platform for the modular sample conditioning, but also the electrical connectivity of the components. This standard proposes a single wire-pair connection to the valves, filters and other modules, and it also allows for a wireless connection. The third phase of the NeSSI brings whole analyzers to the sample conditioning platform. With the miniaturization of optical and other analyzers (Figure 37.22), the promise is that complete instruments will fit the small grid of the platform and will be interchangeable with the same sample access using the same geometry, the same electrical and communication protocols.

37.3.3 PROCESS LIQUID CELLS

Before the sample treatment can be decided, the basic optical arrangement has to be selected. In some online NIR analyzers, the sampling arrangement is fixed; in others, the sampling probe or

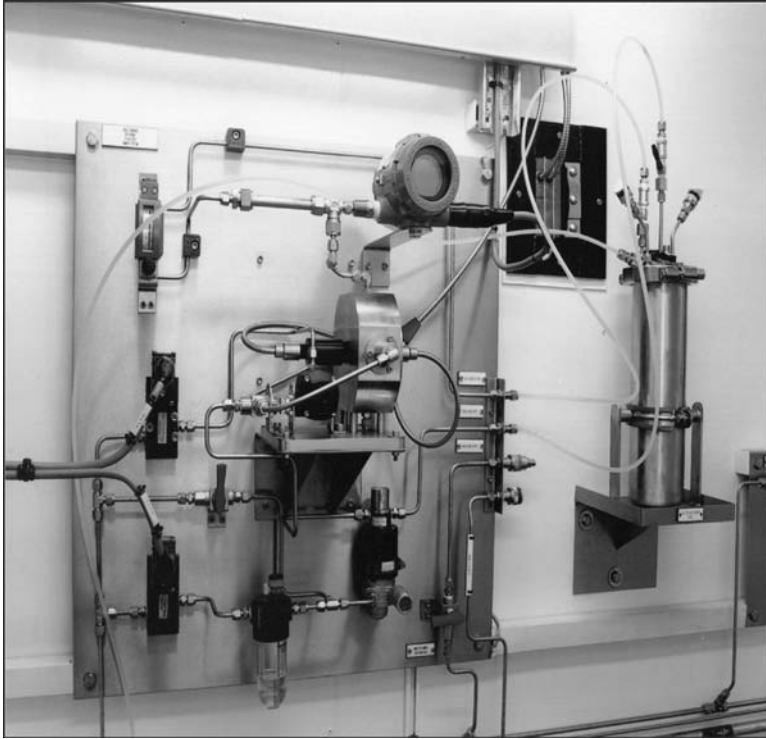


FIGURE 37.18 Sample conditioning system using discrete components. (Courtesy of LT Industries, Gaithersburg, MD.)



FIGURE 37.19 Dedicated sample conditioning system mounted in NEMA enclosure. (Courtesy of LT Industries, Gaithersburg, MD.)

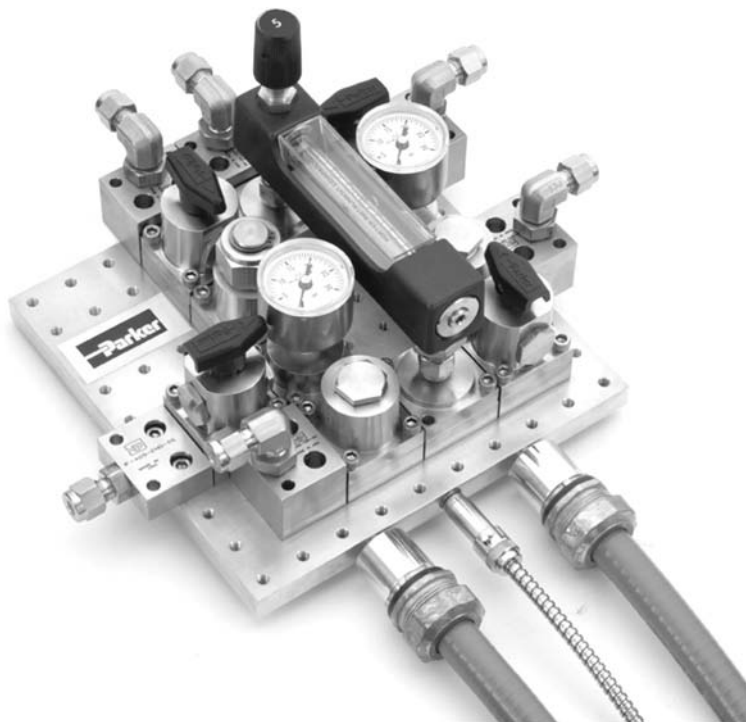


FIGURE 37.20 Sample conditioning system using the ISA/ANSI 76 modular platform. (IntraFlow™, Courtesy of Parker Hannifin, Jacksonville, AL.)



FIGURE 37.21 Sample conditioning system using modular components. Bottom picture is the modular manifold with the standard fluid connections. (Courtesy of Circor International, Inc., Burlington, MA.)



FIGURE 37.22 Optical analyzer as part of the modular sample conditioning platform. (Courtesy of Circor International, Inc., Burlington, MA.)

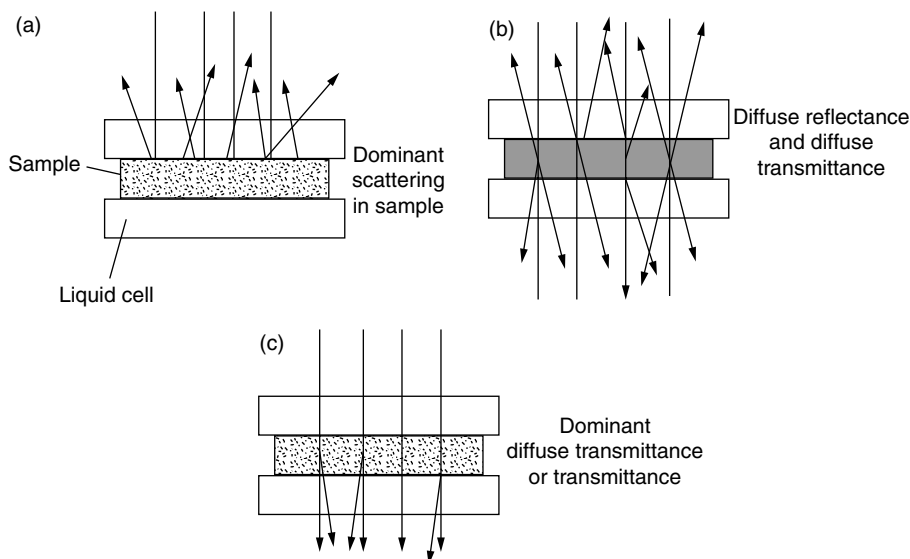


FIGURE 37.23 Liquid sample types containing light-scattering material.

cell is selectable. The first question to be asked is, what type of optical interaction dominates in the actual sample? Figure 37.23 shows three typical cases: (a) fully opaque, (b) moderately opaque, and (c) clear liquids. In the case of the clear liquids, the transmission is dominant; thus the simple transmission arrangement works well if the pathlength is adequately selected.

Liquid streams containing particulate matter are the most difficult to measure well. In the extreme case where the light will not go through the liquid even in a small pathlength, the “classical” NIR reflectance arrangement can be applied. A window should be mounted on the process pipe or vessel and the instrument should illuminate the material through the window. The high concentration of scattering centers will ensure that the reflected portion of the light is enough for the quantitative analysis.

If the amount of scattering is not enough to reflect a large portion of the illuminating light, or if the concentration of the scattering centers varies, a special arrangement, the so-called transreflectance, provides the best quantitative results. Figure 37.24 shows the basic elements of the transreflectance

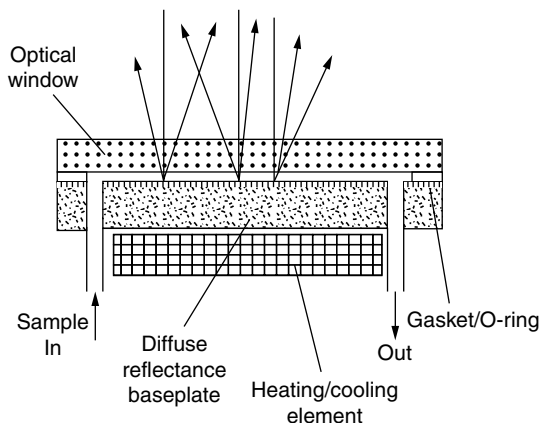


FIGURE 37.24 Transflectance liquid cell.

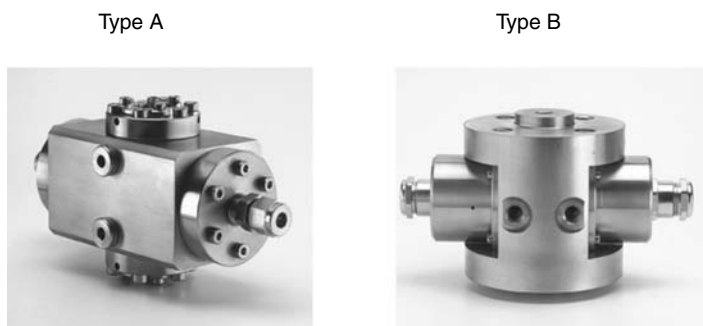


FIGURE 37.25 Process liquid cell for high pressure and high temperature liquids. (Courtesy of Specac Ltd., UK.)

cell. The light first travels through the sample, it is partially absorbed, partially scattered, and partially unchanged. As the light reaches the far side of the liquid cell it has changed its original direction; thus, a wide-angle light collection would be needed in a transmission arrangement. However, a varying portion of the light was also reflected back in a wide angle toward the illumination side, where it could be detected. In the setup shown in the figure, the far side of the cell is made reflective, which will direct the light toward the illumination side in addition to the naturally backscattered NIR radiation. In order to reduce the variance of the optical signal, the reflective surface is made of diffusely scattering material, such as diffuse ceramic, stainless steel, and diffuse gold-coated reflectors. For fully reflecting materials, the collection of the light at a wide angle can be achieved by placing an integrating sphere close to the optical window.

Typical liquid cells for clear process liquids have robust build with the windows strongly sealed to withstand high pressure. The Prospect Type A and Type B flow cells (Specac Ltd., UK) for example, have high pressure ratings of 1500 psi, 280°C temperature limits, chemically resistant Hastelloy bodies, and optional 1- to 10-mm optical pathlengths (Figure 37.25).

A flange-mounted process immersion probe for transmission measurement with deeper sampling is available for sampling reactors or very large diameter pipelines (Figure 37.26).

Integrated liquid analyzers contain the illumination and detection optics along with the liquid cell in one unit [29,30], or the sampling point and the analyzer is connected via a short, large-diameter fiber bundle [31]. The typical integrated liquid analyzer is not mounted into the main process line. Apart from some special cases, the pathlength is too small to allow an appreciable amount of liquid to pass through; therefore, the main stream is sampled in a bypass mode. The bypass allows the



FIGURE 37.26 Process liquid transmission probe. (Courtesy of Hellma GmbH & Co KG.)

insertion of different sample pretreatment devices, conditioning the pressure, temperature, and so forth. The bypass also allows the cleaning of the liquid cell without disruption of the process. The isolation valves on both the sampling and return points can be closed, and the sample loop drained and cleaned.

In addition to these advantages, the bypass mode has obvious shortcomings. The material in the conditioned bypass line may not be the same as the main stream. A longer sampling line and the slow transport will cause a delay in sensing the rapid change of concentration in the process. If the process analyzers are located in a common analyzer house in the plant, the delays can be several minutes, depending on the layout of the plant and the transport rates. One other important factor in a production process is the handling of materials taken out of the process. In most food industries it is not considered good practice to return any sample material once it is removed from the stream. In these processes, the analyzer in a bypass mode will produce waste, albeit a small amount.

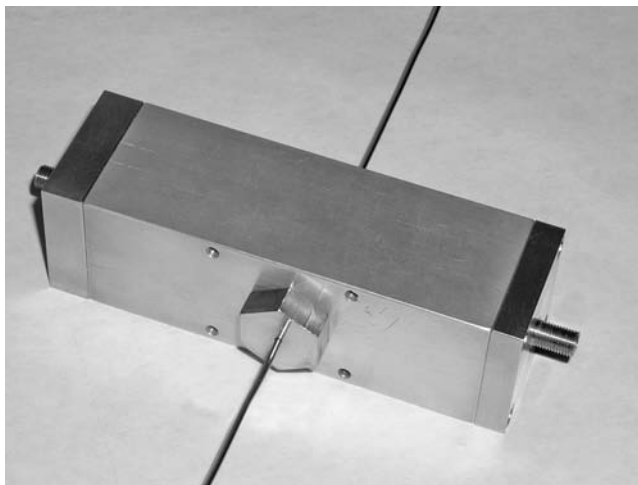


FIGURE 37.27 Liquid cell usable up to 5000 psi. Shown here with its fiber optic adaptor. (Courtesy of Near Infrared Research Corporation, Suffern, NY.)

The main advantage of the integrated liquid analyzers, as opposed to the fiber-optic-based analyzers, is the relatively higher precision of the analysis allowed by the smaller losses in getting the light to the sampling point and by the better light collection. With the development of high performance NIR fiber optic instrumentation, the fiber can serve as a convenient flexible connection of the instrument and sample cell, be it in the process line or off line coupled with sample conditioning. In the NIR region with good available light sources and detectors, the ease of fiber optic coupling between the instrument and the process makes it preferable to moving the sample fluid from the sampling point to the analyzer. The *in situ* fiber-optics-based probes are, in most cases, better than taking sample out of the process with the associated necessary sample conditioning and other complications. One example of this arrangement is a very high-pressure cell developed by the NIR Research Corporation (Suffern, NY). The patented, flow-through high-pressure cell can be used to measure spectra of liquids and gases in the NIR, visible, UV, and mid-IR (to $5.5\ \mu\text{m}$) spectral regions, at pressures up to 5000 psi. The unique design has a sample chamber with a volume of only $2.5\ \mu\text{l}$, while providing high optical efficiency and eliminating dead volume (Figure 37.27).

37.3.4 FIBER-OPTIC-BASED ANALYZERS

The initial reasons for using fiber optics to conduct analyses remote from the sampling point were the hazards associated with the processing of flammable, explosive, toxic, hot, or cold samples (Figure 37.28). Requirements and regulations prohibit in some processes mounting electronics containing, for example, a hot light source and/or high voltage. There are explosion-proof enclosures available, but they are expensive, heavy, and not practical for the packaging of a complete NIR instrument. There are ways around the use of heavy explosion-proof enclosure, for example, purging of the enclosure with an inert gas, and keeping a constant overpressure inside the cabinet. The satisfactory mounting of the optical window and maintaining double safety, in case the window should break, is still not a straightforward engineering job. Consequently, the use of fiber optics to separate electronic functions from the process is increasing in popularity.

The science and technology of fiber optics is a very large and active field. The advances are fueled by the need in communication to transport light pulses with minimal distortion for longer and longer distances. The fiber materials and manufacturing practices, along with the light sources and detectors, are optimized to fulfill the needs of the usually monochromatic light transmission. The

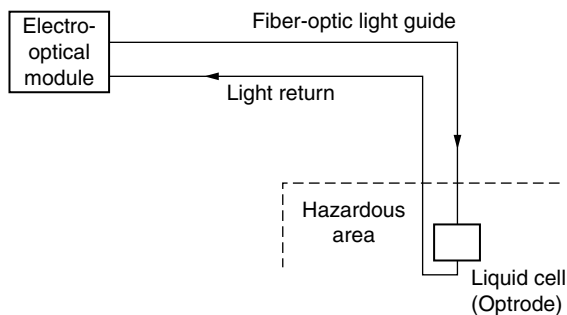


FIGURE 37.28 Separation of the sample cell and the electro-optical part by a fiber-optic guide.

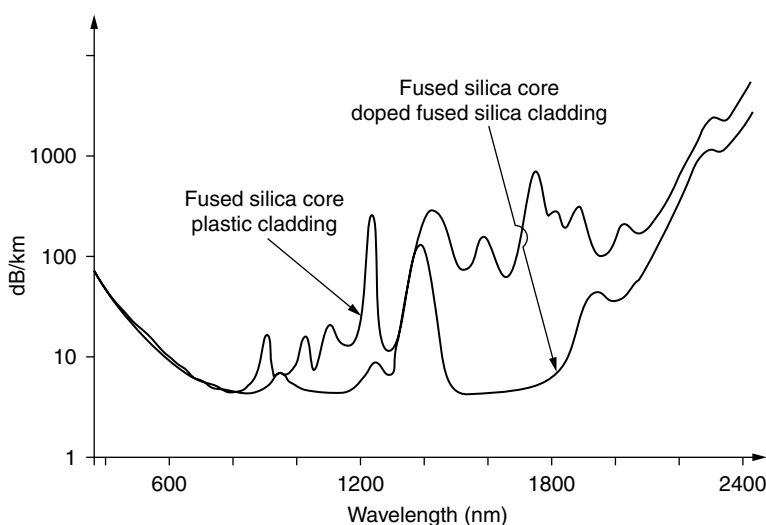


FIGURE 37.29 Attenuation of typical fiber-optic light guides.

application of broadband light transmission has also developed and, although smaller in volume, special fibers are available for the whole optical spectrum, from ultraviolet to mid-IR.

For the purposes of NIR analysis, two basic fiber configurations are currently in use. For longer separation distances between the optical instrument and the sensing point, only *single fibers* are practical. This type of fiber is available from many manufacturers, in thousands of feet length and up to 1000 μm diameter. The other basic type of fiber is a fiber bundle, containing many of the individual fibers, thus increasing the light throughput. Fiber bundles are usually made only a few meters in length, and are very expensive compared to single fibers.

Currently, in the length that fibers are typically applied in analyzers (up to a few hundred feet), the absorption is negligible. Figure 37.29 shows the light attenuation in typical plastic clad and glass clad fibers. The scale of attenuation is dB/km; thus, for example, the worst attenuation (100 dB/km) in a glass-on-glass fiber around 1400 nm is 1:10 in a 100 m fiber. This, of course, is only true when the light is already in the fiber.

The largest light losses occur in coupling the light into the fibers. Typical single-fiber diameters used for carrying light for quantitative purposes are usually 300 to 1000 μm . Most commercially available fiber-optic-based analyzers use incandescent light sources having filaments larger than 1 mm. Even with the most careful design, the numerical aperture (NA) of the fiber and the size of the light source limit the energy that can be coupled into the fiber. Larger diameter fibers are available

in single-fiber packages, but the fibers become increasingly rigid with larger diameters, thus losing the advantages that the fibers can transport the light flexibly into hard-to-reach sampling points. A typical fiber optic-based analyzer contains the source, a grating-based Fourier transform near-infrared (FT-NIR) or tunable filter optical system, the controller, and power supplies. The NEMA 4-rated analyzer housing and the liquid cell is connected with custom-length flexible-steel armored electrical conduit cable. Inside the cable, protected from the industrial environment, resides the 600- or 1000- μm diameter plastic-clad silica fiber. The liquid cells are customized for the application, temperature, pathlength, and material flux through the liquid cell.

The fiber-optic-based liquid analyzers find applications in most chemical or petrochemical processes. The clear liquids are easy applications in the process because the model mixtures can be used in the calibration of the instrument in the laboratory before installation. The best wavelengths for polyols, for example, can be established using general-purpose laboratory analyzers.

Food samples present more challenge to the liquid cell design, as most foods (chocolate syrup, cheese, salad dressings, etc.) contain high concentration of particulates and are therefore opaque. The actual illumination and light collection of the liquid cell determines to a great extent the selection of the best wavelengths. Preliminary feasibility can be proven on other instruments, but the final calibration must be performed using the same optics used for the actual process.

Many improvements in the state of the art of fiber-optic-based liquid analyzers are anticipated. The improvements will target both the fiber-optic cable connecting the analyzer with the sampling point and the analyzer itself.

The fiber illumination optics must be optimized to achieve the lowest noise and drift in the overall system. This criterion is met if the light energy is collected onto the fiber with the cone angle allowed by the NA of the fiber. Some overfilling in cone angle and some overfilling in illuminated area, larger than the cross-section of the fiber, was postulated to be the best design by the late Thomas Hirschfeld. The overfilling somewhat reduces the sensitivity of the optical system to induced variations of the transmitted light intensity. Vibration, optical tolerances, and changes in index of refraction due to temperature fluctuations are some of the factors that alter the intensity of the transmitted light.

Several companies offer different types of turnkey fiber optic-based NIR analyzers. Some examples are

- *FT-NIR based:* Thermo Electron (Waltham, MA), Bruker Instruments (Ettlingen, Germany)
- *Dispersive/Diode Array based:* Foss (Eden Prairie, MN), LT Industries (Gaithersburg, MD), BWTek (Newark, DE), Analytical Spectral Devices, Inc. (ASD, Boulder CO), and Guided Wave (Rancho Cordova, CA)
- *Acousto-optic tunable filter (AOTF)-based:* Brimrose (Baltimore, MD), Infrared Fiber Systems (Silver Springs, MD).

Of all the currently available technologies, there are some potential new technologies that are inherently faster than the previous generation of rotating filter and scanning grating-based process analyzers, because the wavelength is selected electronically. Detector arrays coupled with a grating allow very fast, and fully parallel measurement at all wavelengths without any moving parts. All detector elements are exposed at the same time to the light resolved by the grating. The readout is a very fast serial or, in some detector arrays, coupling individual amplifiers to each detector element, a fully parallel electronic task (Figure 37.30). There are various detector arrays available in the NIR, such as silicon (sensitive in the 200–1100 nm region), indium–gallium–arsenide (InGaAs in the 900–1700 nm region), extended InGaAs (1100–2300 nm region), lead-sulfide (PbS, in the 1000–3000 nm region), and NIR enhanced mercury–cadmium–telluride (MCT, in the 1500–4000 nm region). The MCT detectors are new in the NIR range, with a very good potential for rugged, sensitive, and wide wavelength region sensing. Other extended wavelengths are also available with λ_{peak} at 2.8, 4.8, or 7.5 μm .

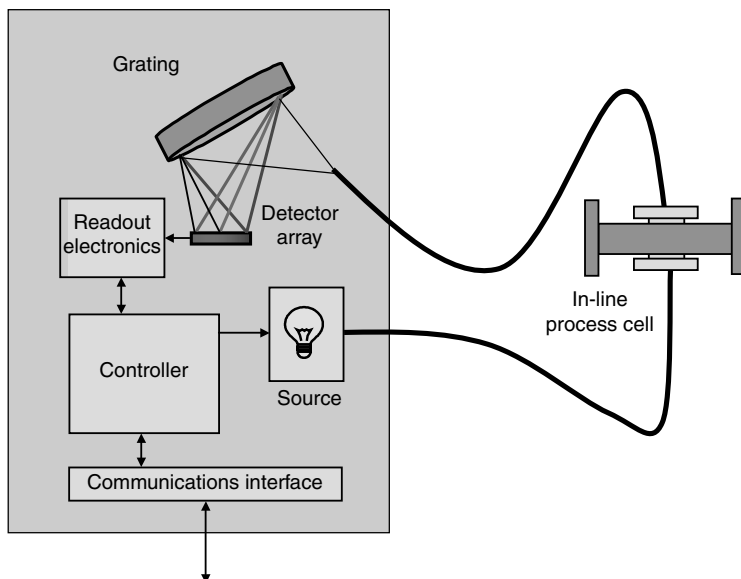


FIGURE 37.30 Block diagram of fiber-optic based detector array measurement.

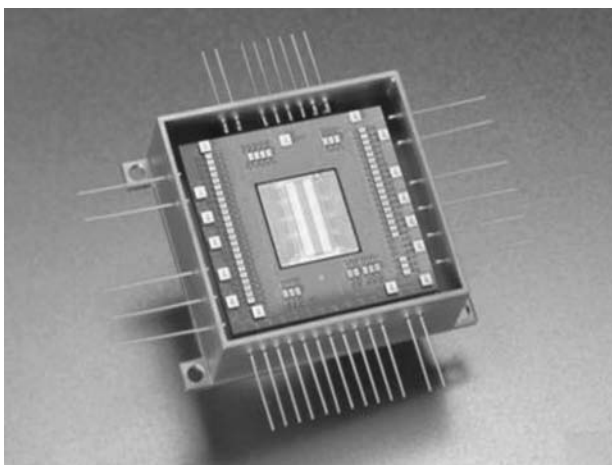


FIGURE 37.31 Near-infrared mercury-cadmium-telluride (MCT) detector array. (Courtesy of Optical Instrument Centre, VTT, Oulu, Finland.)

The development of the InGaAs detector array technology made it possible to expand the wavelength further into the NIR. Some of the recently developed detector materials, such as the NIR-MCT, require cooling, but the simplicity, sensitivity, and ruggedness of the optical system make this approach attractive; thus, new development is expected in this direction (Figure 37.31).

FT-NIR are becoming available in process formats. The instrument-to-instrument wavelength repeatability of FT-NIR instruments is an important factor in calibration transfer among instruments allowing less calibration work when multiple sensing points are required. The Matrix-F FT-NIR instrument is an example of the sealed and rugged process instruments with multiple fiber access points [32]. The fiber multiplexer is able to select a fiber pair representing one sampling location to which a dedicated fiber probe is connected (Figure 37.32).

Another promising technology is the AOTF. The principle of operation of AOTF has been known since the late 1960s. As shown in Figure 37.33, the broadband light is directed onto a TeO_2 crystal.

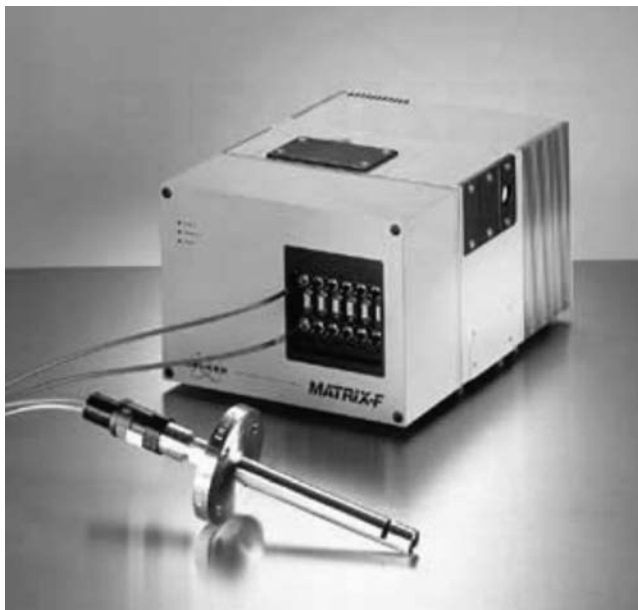


FIGURE 37.32 Matrix-F FT-NIR process analyzer with six pairs of multiplexed fiber optic ports. (Courtesy of Bruker Optics, Billerica, MA.)

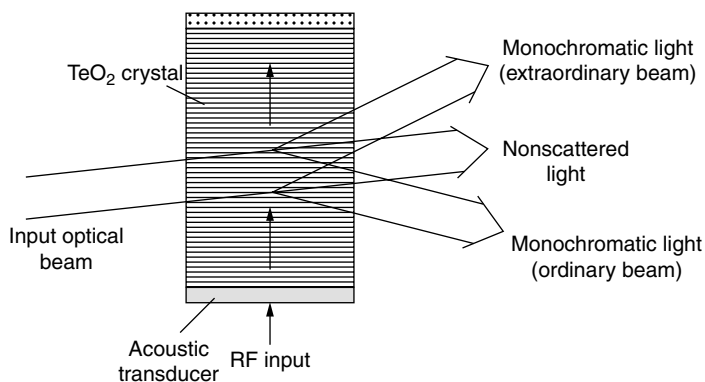


FIGURE 37.33 Principle of operation of acousto-optic tunable filters.

The TeO₂ material has virtually no absorption from the visible spectrum to about 5 μm , and it has good acousto-optic properties from the visible through the NIR region. On one surface of the oriented crystal, an acoustic transducer is bonded. The acoustic transducer is a piezoelectric material, in this case LiNbO₃ driven by 1 to 4 W of radio frequency (RF) coupled into the transducer. The high-frequency (30 to 200 MHz) acoustic waves induce index of refraction waves in the acousto-optical material. The waves travel through the crystal very quickly. Typically within 20 to 30 μs , the acoustic waves “fill” the crystal, interacting with the broadband light traveling through the crystal. The angles of the crystal axis, the relative angles of the broad-band light, and the acoustic wave are defined in the crystal design to optimize the light-acoustic wave interaction. The result of the interaction is the splitting of the light into three beams. The center beam is the unaltered white light traveling through the crystal. The two new beams generated by the acoustically excited crystal are monochromatic and orthogonally polarized. These beams can be used as monochromatic light sources for analytical purposes.

The main advantage of the AOTF optics is that the wavelength is electronically selected. The electronic wavelength selection allows a very high-duty cycle, because *wavelength selection* requires only that the acoustic wave with the new frequency fill the crystal, which takes a few microseconds. In comparison with fast-scanning mechanical grating instruments, the advantage is not only that the scanning rate is orders of magnitude faster, but also that *the wavelength access is random*. If only four or five selected wavelengths are required for the concentration equation, the AOTF instrument is able to select those, and is not confined to accessing all wavelengths serially, as is the case with fast-grating monochromators, diode array systems, or FT-NIR.

Besides the speed and efficiency of wavelength selection, the acousto-optical tunable spectrometers (AOTS) are much smaller than grating monochromators with equal resolution. In a properly engineered design the *long-term wavelength repeatability* also surpasses that of the grating monochromators.

The first practical devices were used in the 1970s for dedicated optical purposes. The first report on demonstrating the use of a complete computerized spectrophotometer for NIR analytical purposes based on AOTF was presented in 1986 [33,34]. Bran and Luebbe demonstrated an AOTF-based NIR liquid analyzer in 1990. Other companies, such as Brimrose, Baltimore, MD (Luminar Model 3010/4010) and Infrared Fiber Systems, Silver Springs, Maryland), are also offering AOTF-based process spectrometer systems [35].

The high performance (better than 30 μ Abs standard deviation among repeated scans taken less than a second each) of the AOTF-based system of the Bran and Luebbe instrument was achieved by the proprietary dark-field dual-beam optical module, the special high-speed RF synthesizers, the low-noise high-speed detector-preamplifier module, and the efficient fiber-optic coupling [36]. Figure 37.34 shows a typical single-scan spectrum of chloroform. The time spent to collect any one point in the spectrum was approximately 100 μ s.

37.4 CONTACT ANALYSIS OF SOLIDS

37.4.1 SAMPLING PARAMETERS

In a wide variety of agricultural, food, chemical, and pharmaceutical industries, there is a growing need for rapid online analysis of several parameters of particulate materials. In comparison

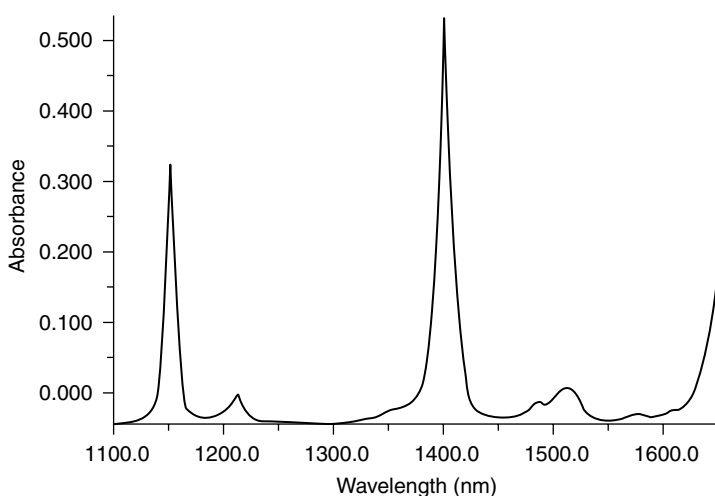


FIGURE 37.34 NIR spectrum of chloroform in 10-mm-pathlength liquid cell. The spectrum was recorded with an InfraAlyzer AOTS.

with process liquids, it is much more difficult to handle solid samples, raw materials, intermediate products, or final products by continuous analytical methods. In addition, solid materials, crystalline powders, pelletized plastics, and animal feed mixtures are typically inhomogeneous. Samples of these materials are consequently, physically and chemically inhomogeneous.

Concentration inhomogeneities may be caused by mixing different materials or the same type of materials with different concentrations. If the inhomogeneities are in a macroscopic scale comparable to the sample cuvette size or the volume of the optical cell in which the measurement occurs, the local concentration variation will affect the measurement. Beyond the concentration variation of the sample, the optical physical characteristics of the sample also have a profound effect on the result; therefore, special sampling techniques must be developed.

NIR analysis is directed to the determination of bulk properties and concentrations of the sample. In order to ensure precision of analysis, a sufficient number of particles must be present in the sample cell. Hirschfeld [37] discussed the relationship of measurement error as a function of sample area geometry and the average diameter of the particles. One effect of large particle size in a solid sampling diffuse reflectance cell is that the light penetration changes, distorting the spectrum. This is a known phenomenon in the visible-NIR [38,39],

$$\Delta T_n = \frac{d}{D} \sqrt{n\varepsilon} \quad (37.8)$$

where ΔT_n is the transmittance error, d is the particle size, D is the area of illuminated sample, n is the light penetration depth expressed in the number of particle layers in which the intensity (I_0) of the incident light diminishes to I_0/e , and ε is the transmittance variance between particles. The stochastic noise due to particle nonuniformity and the presence of a finite number of particles in the field is discussed under two different sets of optimization conditions. One optimization was carried out with the detector size varied in the other calculation the detector size was fixed. In both treatments, the calculations with actual particle size diameters and sample area show that the sample nonuniformity is a dominant noise source, even for apparently small particles and fairly uniform particulates. The optimum conditions with sample diameters of several hundred micrometers call for large sample areas. In an example, assuming variable-size detector, $d = 500 \mu\text{m}$ average particle size, three-layer light penetration, $n = 3$, and 5% transmittance variance between particles. The optimum sample diameter is 2.79 cm. The optimum diameter with fixed detector size, on the other hand, is 6.28 cm.

Ideal contact solid sampling means presenting the solid sample for the optical analysis close to the optical analyzer contacting the optical window. However, transporting powdered solids through a sample cell is much more difficult than liquids. Liquids fill the volume of the cell completely; bubbles, if any, rise to the top of the cell leaving a continuous fluid medium at the bottom part of the cell to be measured. Solid powdered materials, on the other hand, flow only if a certain flow angle is maintained to overcome the cohesion of the particles. This angle is characteristic of the respective materials. Ultrasonic or other mechanical devices are needed, for example, to empty storage silos, if the cohesion of the particle is larger due to increased moisture content. Similarly, ultrasonic agitation can be used to loosen the particles of solids, and pressurized air can enhance the transport through pipes and chutes.

The optical concentration analysis is based on the light interacting with the material of a certain optical density in a defined reproducible pathlength. In liquids, a fixed pathlength is maintained that along with the relatively constant physical density of the liquid allows uniform sampling. It is, however, extremely difficult to keep both the density and the optical pathlength constant in solids. Powdered solids are compressible under normal pressure, unlike liquids. The filling factor is known to influence the light scattering and light penetration depth considerably. In an experiment, whole-grain wheat was poured gently into a sample holder of a Trebor-90 NIR transmission analyzer. This way, it occupies a much larger volume because the individual kernels are randomly oriented. When agitated by ultrasonic vibration or by a Vortex test-tube mixer, up to 15% settling can be observed [40]. This naturally increases the average material density in the optical path, and results in much

lower transmittance readings at all wavelengths. The settling effect was different for different types of wheat due to the different average shapes and sizes of the varieties.

In the laboratory NIR analysis, the manufacturers have developed a string of solid sample cuvettes. The filling factor is known to influence the light scattering and light penetration depth considerably in all solids. The reproducible material density is achieved by rubber cushion or spring mechanisms. Even by the application of the same force, reproducible cuvette filling can only be achieved by reproducible cuvette-loading practices. If the solid cup is filled properly, the powder fills the space behind the optical window. Improper packing can result in air gaps and crevices in the powder at the optical window. These standardized conditions have to be ensured by the online sampling devices too.

37.4.2 SOLID (POWDER) ANALYZERS

Contact solid analyzers can be used if the solid materials can be made to come in contact with the optical window, which is ideally an integral part of the process analyzer. The sample is analyzed through the optical window. This has the distinct advantage that the sample plane is at an optimal position for the measurement, the sample is protected from the ambient light, and good instrumental precision can be achieved. If the sample material does not scratch the window material, does not form irreversible deposits on the window, and can be handled to move in and out of the sample cell, a measurement precision close to that of the laboratory-type NIR analyzer can be achieved. The use of hard, abrasion-resistant materials (e.g., sapphire) for the window is suggested because even the softest benign sample (e.g., flour) may contain microscopic very hard particles, which in a continuous long operation could scratch softer glass. The formation of microscopic scratches causes hard-to-detect drift in the concentration output.

One of the first commercial online contact NIR instruments was a version of the PerCon Inframatic NIR diffuse reflectance laboratory analyzer [41]. The instrument was used in the flour industry to measure protein, ash, and moisture content on a continuous basis [42]. In the online measurement six standard filters (2310, 2230, 2180, 2100, 1940, and 1680 nm) were used. The flour stream in a bypass chute was interrupted, with the flour compacted against the optical window. The instrument was placed in a bypass chute from the main flow. Flour remaining after the optical measurement was returned to the main flow.

A NIR solid sampling system for animal feedstuffs, flour, milk powder, egg, and soy products, cement, and lime is described in British Patent 2,150,917B. In order to achieve reliable representative sampling, the continuous stream of material is allowed to fall freely and a sample trap, positioned in the stream on two sides of the optical window, is closed intermittently. The trapped material is compacted to allow uniform density for the time of the optical measurement. After the measurement, the trap opens on the bottom, and the aliquot material returns to the main stream. Compaction of the material can be achieved in different sampling or an Archimedian screw, according to the above British Patent (Figure 37.35), or pneumatically via a plunger (Figure 37.36). In the latter system, the sample removal can be enhanced by compressed air between consecutive measurements to clean the optical window.

Sampling systems can be constructed with a paddle-wheel-type mechanism moving the sample in front of the sample window or by selecting a process location where the sample is continuously moving but still assuring constant coverage of the optical window. This type of sampling was implemented in the flour mill of Kansas State University [43,44].

With the new PAT initiative in the pharmaceutical industry, a great deal of effort is directed towards monitoring the various unit processing steps, such as incoming material identification and establishing the endpoint of the blending of solid ingredients.

Laboratory NIR analyzer platforms have been modified, enclosed in process hardened housing, and used extensively for material identifications in the pharmaceutical [45,46] and the nutraceutical industries [47].

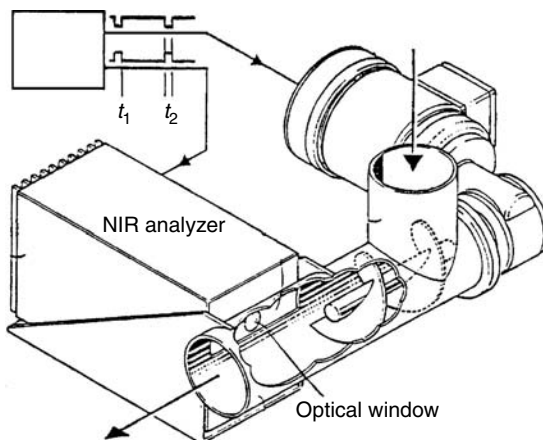


FIGURE 37.35 Screw conveyor-type sampling system. (After British Pat. 2,142,721A.)

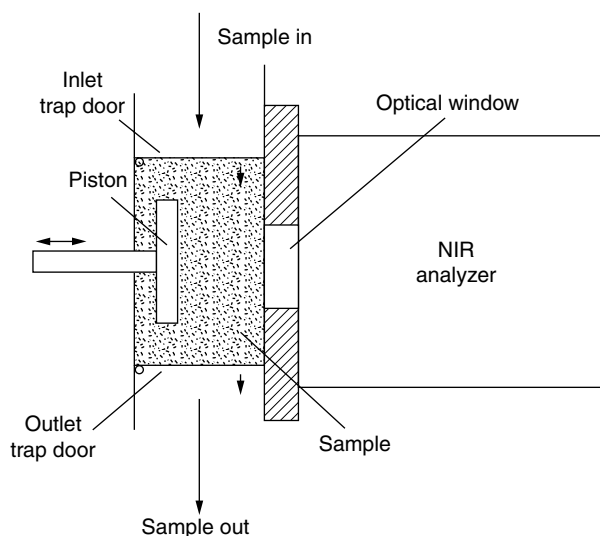


FIGURE 37.36 Sampling system to ensure uniform compression of sample.

Blend monitoring and end-point detection with various analytical techniques has become a large, active field because all tableting operations are preceded by blending of the active pharmaceutical ingredient (API) with binders, fillers, disintegrants, flow-enhancers, and other excipients. NIR is one of the most often used techniques [48–50] with many companies offering dedicated blend monitors.

The blend analyzers must be capable of operating with the blender in motion. Most analyzers offer battery powered options and wireless communication capabilities. MEMS tunable Fabry-Perot-based blend monitors are offered by Thermo Electron Corporation (Waltham, MA, Figure 37.37), Expo Technologies, Inc. (Chagrin Falls, OH), and Sentronic (Dresden, Germany). MEMS tunable Fabry-Perot modules are electronically tuned interferometers (Figure 37.38). Super-luminant LED sources provide NIR radiation with much higher intensity than incandescent quartz halogen lamps, which offsets the throughput allowed by the small interferometer and the fiber coupling [51]. The Fabry-Perot wavelength selection allows optical resolution better than 1 nm, and tuning in the 1350 to 1970 nm range. It is anticipated that wider wavelength range devices will be produced by combining multiple MEMS units and adding new LED sources.



FIGURE 37.37 AntarisTarget blend analyzer. (Courtesy of Thermo Electron Corporation, Waltham, MA.)

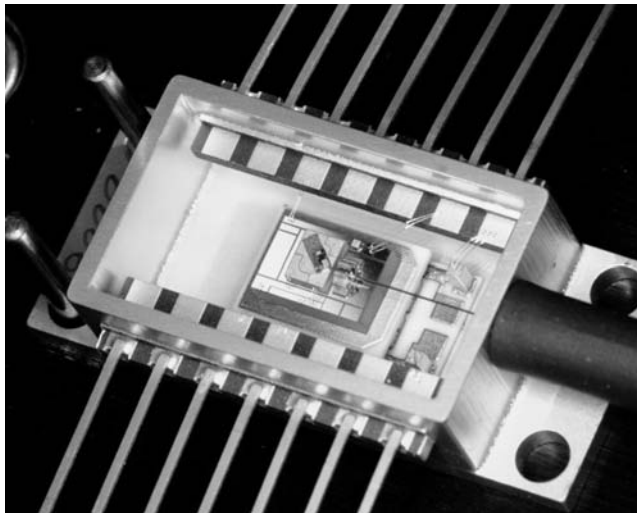


FIGURE 37.38 MEMS tunable Fabry-Perot module with fiber optic output. (Courtesy of Axsum, Billerica, MA.)

Stability of the light output is a requirement to be able to monitor small spectral changes. Source intensity and wavelength has to be stabilized internally in the MEMS Fabry-Perot modules (Figure 37.39).

Other blend analyzers are based on InGaAs arrays: CDI Pharma (South Bend, IN), Zeiss (Jena, Germany), and on AOTF: Brimrose (Baltimore, MD). Similar instrumentation can be used to monitor granulation and drying and other unit processes. Moisture measurement during granulation was monitored with a fixed wavelength sensor (WET-EYE, Fuji Paudal Co., Osaka, Japan) [52]. The mixing ratio of lactose monohydrate and cornstarch was found to affect the moisture calibration of the sensor [53]. A system with four fixed wavelengths, developed by the Optical Instrument Centre,

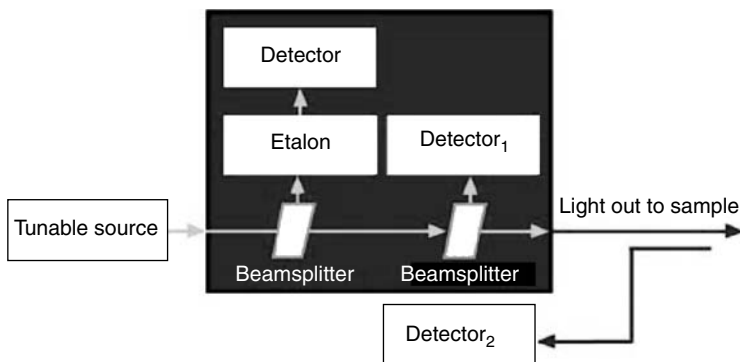


FIGURE 37.39 MEMS tunable Fabry-Perot module. (Courtesy of Axsum, Billerica, MA.)

VTT, Finland, was reported for the monitoring of fluidized bed granulation [54]. A fluid-bed monitor was developed with sapphire optics that separate the material and analyzer (The PH710 by NDC Infrared Engineering, Irwindale, CA).

Freeze drying, a widely used process step in the biopharmaceutical industry, was studied by Derksen et al. [55]. Using NIR spectroscopy, residual moisture can be measured not only during the freeze-drying, but also nondestructively through the glass in the sealed ampoules containing the dosage unit of the protein drugs [56].

A process diffuse reflectance probe that works in contact with the sample, is available from Hellma GmbH & Co KG. The Model 661.761.NIR probe has seven 200 μm diameter illumination fibers illuminating the sample at a 30° angle, and one 600 μm diameter core pickup fiber at the center of the illumination fibers. The DN 25/PN40 standard flange mounted probe withstands a temperature ranging from 5 to 140°C and 10 bar pressure. Chemical resistance can be enhanced with the Hastelloy™ C-22 body (Figure 37.40).

37.5 NONCONTACT ANALYSIS

Analysis in the NIR range is nondestructive; therefore, it would be ideal, if the analysis could be done without even coming in contact with the process sample. As can be seen from an earlier chapter, pulling samples from liquids or solids is fraught with dangers for the process, requires costly sampling devices, and, most expensive of all, requires constant maintenance and supervision.

Noncontact analysis is performed using an optical head that does not touch the sample. However, the sample positioning is not as reproducible as in contact sampling, and the analyzer is affected by several factors associated with the sampling and process environments.

37.5.1 FACTORS INFLUENCING NONCONTACT ANALYSIS

The analyzer signal output, which should ideally be the concentration only, will be affected by

1. Local concentration variations
2. Optical variations
3. Sample — sensor distance variations
4. Sample speed variation.

Other process conditions influencing the noncontact analysis are

5. External (ambient) light
6. Temperature variation of the sample



FIGURE 37.40 Process NIR diffuse reflectance probe. (Courtesy of Hellma GmbH & Co KG.)

7. Ambient temperature changes
8. Mechanical vibrations
9. Power-line fluctuations
10. Dust
11. Corrosive vapors (steam, water vapor).

Figure 37.41 shows the schematic sampling arrangement where the sample is transported on a conveyor belt. In a real sample, even if the overall concentration is constant, the individual particles of the sample could contain different concentrations. The changes in the concentration alter the individual optical readings at different wavelengths.

The particle size of the sample changes with the source of material, or with the adjustment or wear of the process grinders and mixers, or with changing moisture content. All of these will change the optical readings in a nonlinear fashion. Changing particle size affects the diffuse light scattering at practically all wavelengths, but not necessarily to the same degree.

The sample-to-sensor distance can also change in a manufacturing environment. The control devices in the overall process are keeping the temperature, pressure, or other process parameters constant in a closed loop. One of the process variables commonly controlled to achieve, for example, the targeted moisture value is the material transport rate. The material stream can vary in a wide range in an automated process. The change of material flux affects the cross-section of the solid material transported and thus the level of the material. If the sensor is over the sample flow, the increased flux results in a reduced sample-to-sensor distance. Sample distance variations change the optical readout; thus the manufacturers of noncontact analyzers specify the limits within which the concentration variation is still acceptable. This is usually about ± 1 to ± 2 in. from the nominal sample distance. The different noncontact optical designs are usually optimized to reduce the unwanted effects of the sample distance variation.

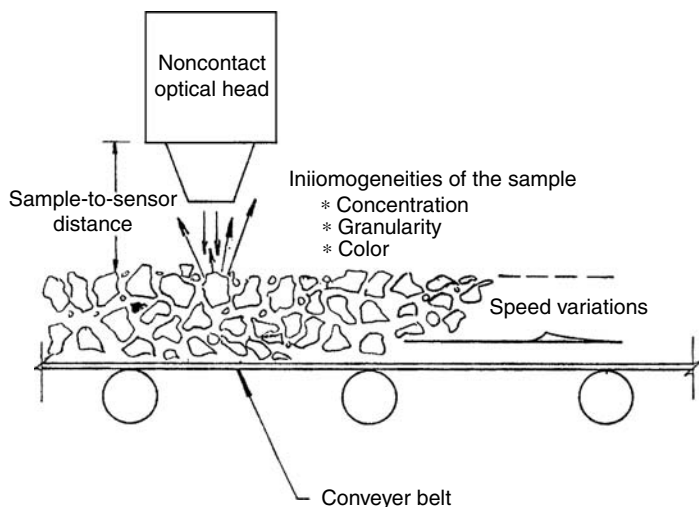


FIGURE 37.41 Noncontact optical arrangement.

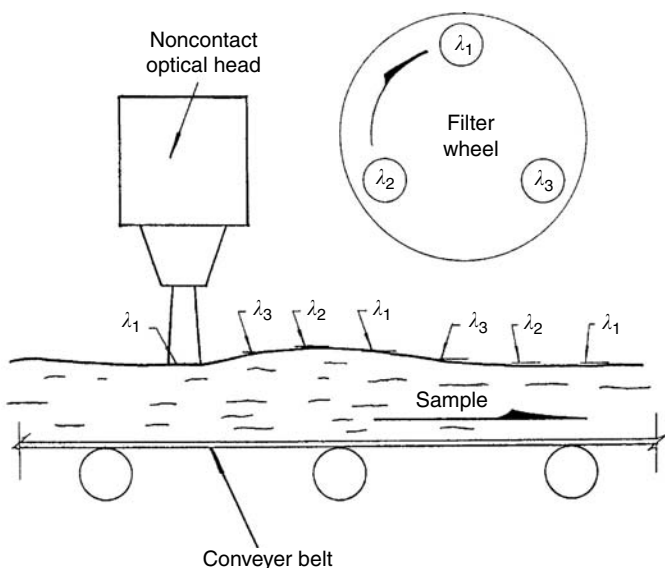


FIGURE 37.42 Analysis of moving sample with multifilter noncontact optics.

37.5.2 SAMPLE SPEED VARIATION

Sample speed variations are known to interact with multiwavelength analysis. Figure 37.42 outlines a typical noncontact analyzer with a rotating interference filter wheel, representing any optical analyzer that projects one selected wavelength of light on the sample at a time. The light with the respective wavelength is flashed onto the sample and the reflected signal is detected. As the filter wheel is rotated, a small portion of the sample is examined with the light at single wavelengths. The next light with the different wavelength is flashed at a different portion of the sample. The consecutive readings shown in Figure 37.43 are combined in an analog or digital fashion to provide the concentration readout. If the values at the same wavelength, or if the computed concentrations are different, the values can be averaged to reduce the error associated with the measurement.

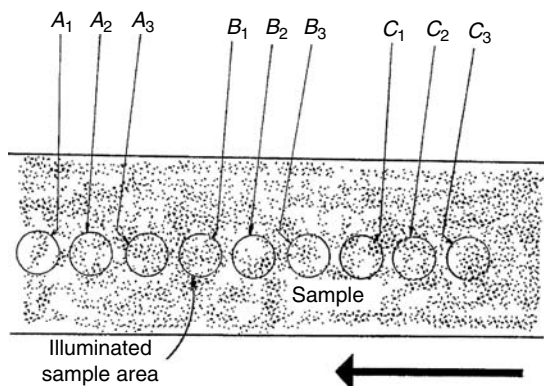


FIGURE 37.43 Illumination of the sample with rotating multifilter analyzer.

Averaging is known to reduce the error of the overall result by a factor of square root of N , if N readings are averaged. If the error associated with the individual readings is known, the number of required averages can be calculated.

$$\varepsilon = \frac{\sigma}{\sqrt{N}} \quad (37.9)$$

where N is the necessary number of independent readings to achieve ε error, σ is the standard deviation of the individual readings. This formula is, however, valid only in ideal cases, where the analyzer is both infinitely faster than the sample changes and the consecutive data points can be considered statistically independent. The theoretical treatment of random and correlated signals has an extensive literature, because the treatment of random signals and noise is at the heart of many physical and electronic principles, such as astrophysics, radar, and other communication technologies [57].

This measurement on continuously moving samples in statistical terms is a sampling of continuous random variables. In the simplest case, the sample is moving and is not homogeneous, but the targeted average concentration is the same. The detector signals can statistically be considered stationary random processes. Equation (37.9) is only valid for readings that are statistically independent. NIR analysis in real time is much more complicated because of the fine system of correlations between the data points. There is a measurable, finite correlation between the observed values at different wavelengths on the same or very close sample spots: $A_1 - B_1 - C_1, A_2 - B_2 - C_2$, and so forth (Figure 37.43). Correlation is found between consecutive readings: $A_1 - A_2 - A_3, B_1 - B_2 - B_3$, and so forth, and between the calculated consecutive concentrations (Y_1, Y_2), and so forth.

In a single-variable case, let x be a stationary random process variable with a mean m and a finite variance σ .

$$E(X_n) = m_X \quad (37.10)$$

where $E(x)$ is the first moment that is the average of the process variables. The real stationary value of this process variable is estimated by averages calculated for shorter time intervals.

$$M = \frac{1}{N} \sum_{n=1}^N X_n \quad (37.11)$$

The statistical average of this sample mean is

$$E(M) = \frac{1}{N} \sum_{n=1}^N E(X_n) = m_X \quad (37.12)$$

The statistical average of the sample mean is equal to the mean value of the random process variable. This means that averaging the local averages gives an “unbiased estimator” of the true process variable in a stationary process. If we allow the number of measurements to increase without limit:

$$\lim_{N \rightarrow \infty} \sigma^2(M) = 0 \quad (37.13)$$

means that as the number of measurement increases, the sample mean converges to the desired mean, called the limit in the mean. The sample mean becomes a better and better estimator as the number of measurements increases. However, the rate with which this convergence occurs is very much dependent on the nature of the individual readings. The variance of the sample mean is

$$\sigma^2(M) = \left[\frac{1}{N^2} \sum_{n=1}^N \sum_{m=1}^N E(X_n X_m) \right] - m_x^2 \quad (37.14)$$

where $E(x_n x_m)$ is the correlation term between the n th and m th data points.

This describes an ideal case, where the consecutive readings are uncorrelated:

$$\sigma^2(M) = \frac{\sigma_x^2}{N} \quad (37.15)$$

The other extreme case is where the samples are fully correlated:

$$E(x_n x_m) = E(x^2) \quad (37.16)$$

for all m and n values. If the samples are highly correlated, the variance of the sample mean approximates the variance of the random variable itself in the extreme case for any number of samples. The effect of this is that any number of measurements will not improve the estimation of the desired mean compared to the single measurement.

In a nonstationary process, compositional changes should be detected, this being the goal of the continuous analysis. Homogeneous samples, films, and webs for example, move at very rapid speeds; thus, defects produce very fast transients. Even slow-moving materials produce fast changes in composition if they contain small inhomogeneities, different size particles, or sudden color changes. The NIR process analyzer is a dynamic system in which the output is a function of the spectrum of the sample aliquot, as well as the rate of change of the spectrum. Measuring the spectrum, or spectral data points of moving and changing samples, is much more complex than that of the stationary sample [58]. The response of the optical analyzer to transient optical or concentration change is dependent on the frequency response of the detector and signal processing, but even more so on the basic optical design of the analyzer. Figure 37.44 shows the two basic types of optical systems. The previously described multifilter optical head probes the sample with monochromatic light packets, one at a time. Because the sample moves in between the sampling events, the readings at different wavelengths cannot be considered to “belong together.” Since this incoherence happens over time, the effect is called temporal incoherence. Figure 37.44B has fixed detectors that measure at the same time, however, they receive images of different parts of the sample. This is called spatial incoherence, and again causes a distortion in the measurement. In stationary processes, the readings of both the temporal and spatially incoherent readings converge, and the correct reading is produced at the end of the averaging cycle, as shown above. The rate of convergence to the target value, however, is dependent on how closely the individual wavelength readings represent the true values, and how many readings can be averaged. In the practice of analysis of moving samples, the slow mechanical scanners have been replaced by systems with better temporal coherence and with spectral scans at electronic speed, thereby yielding many more data points to average.

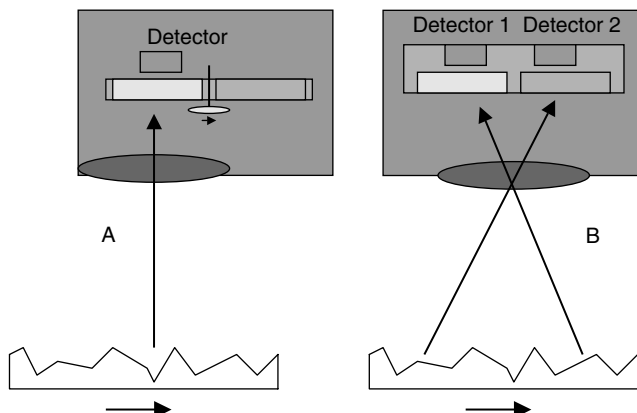


FIGURE 37.44 Two basic types of optical sampling incoherence: A — Temporal, B — Spatial.

37.5.3 NONCONTACT NIR ANALYZERS

Noncontact NIR analyzers were known and widely used since the early 1970s. The first use of these analyzers was in measuring moisture content of samples moving on conveyers. Several two- or three-color interference filter-based instruments are still manufactured and sold for applications in the inorganic, tobacco, food, chemical, plastic, and other industries.

Noncontact instruments have some common features. The optical module illuminates the sample to be measured from 5 to 12 in. NEMA-4 or an equivalent-rated instrument protection is provided for both the optical and the electronic modules.

There are several thousand NIR moisture sensors installed in different industries, but due to the competitive and proprietary nature of the processes, there have been very few reports in the literature about successful, specific applications. From among the reported applications, a Moisture Register NIR analyzer was used to look at the moisture content of crackers moving at a speed of 100 ft/min [59]. Other early examples were paper moving at 50 ft/min web speed and phosphate ore-drying process monitoring [60]. A comparative study of different noncontact instruments was reported to measure cake moisture in wastewater treatment [61].

Although moisture provides a large and broad analytical band at 1.94 and 1.4 μm for noncontact analysis, there have been several successful applications to measure constituents other than moisture. Noncontact moisture analyzers are offered with optional filter wheels sensitive to other constituents. Multifilter analyzers are capable of detecting complex, weakly absorbing constituents, such as nicotine and sugars in tobacco. The Model TM710 of NDC Infrared Engineering is optimized to measure moisture, sugars, and nicotine in tobacco leaf processing [62]. Versions of the original noncontact instrument designs were used in webs and films to measure thickness and composition. A 15-filter plastic film analyzer was reported [63].

The noncontact arrangement is often the most suitable in processes with solid materials. The material does not have to be diverted, ground, and wasted after the measurement. Meat composition analysis using a diode array-based analyzer has been reported [64]. In this application, the meat was moved continuously on a conveyor belt. An important consideration for the selection of process analyzer sampling type is how the sanitary sampling point can be designed. In the noncontact arrangement, the food or pharmaceutical material is transported without direct contact by the analyzer and without diverting a portion of the process stream, thereby meeting all sanitary requirements.

A very cost-effective and fast measurement can be achieved by the no-moving-part, multidetector system comprising a set of interference filters that are placed in front of the individual detectors in the array (Figure 37.45).

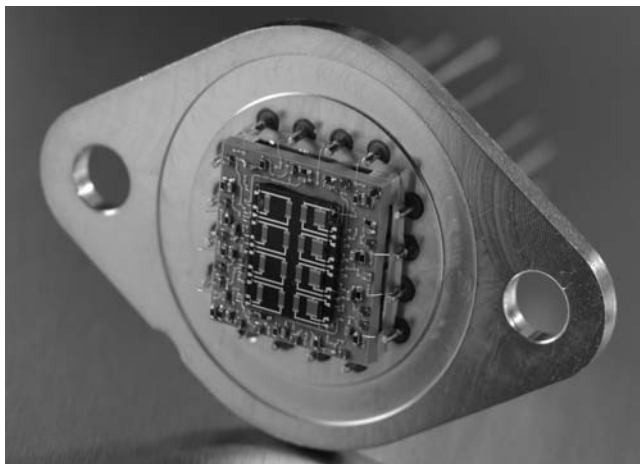


FIGURE 37.45 4×4 PbS detector array module. (Courtesy of Optical Instrument Centre, VTT, Finland.)

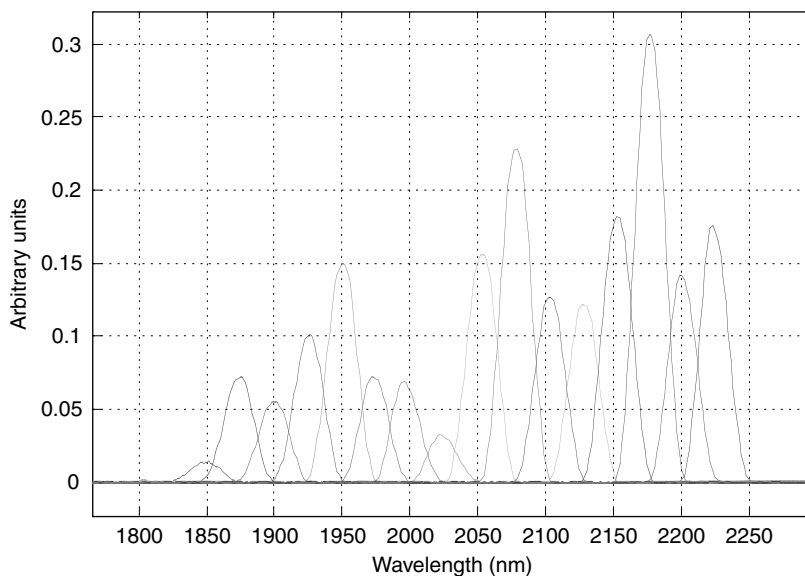


FIGURE 37.46 Sixteen wavelength LED module for very high speed wavelength switching. (Courtesy of Optical Instrument Centre, VTT, Finland.)

Another fast and efficient system uses NIR LEDs with a grating, narrowing the emitted light. Figure 37.46 shows the selected sixteen wavelengths that can be turned on or off, multiplexed at electronic speeds without moving components.

A noncontact FT-NIR spectrometer with several light sources has also been developed. The system has light collection capabilities for the sample placed between 17 and 50 cm from the optical head. Moving samples that may contain large particles, are measured by an infrared beam projected at approximately 25 mm [65]. The instrument window of the analyzer has a built-in wiper to keep the window clean in demanding process environments.

Certain industrial monitoring requires more than a single noncontact measurement. In order to characterize paper, for example, up to three optical heads can be directed at one measurement point, measuring in parallel the transmission, the specular reflectance (gloss), and the diffuse reflectance



FIGURE 37.47 Multihead noncontact web measurement at KCL pilot coater in Finland. (Courtesy of VTT Optical Instrument Centre, Finland.)

of the web (Figure 37.47). Another extension of noncontact analysis is the measurement of the full width of the web. This can be done by moving the optical head across the width of the web, or by parallel and multiple measurement points. Yet another extension of the noncontact technology is the measurement of both sides of a web.

Near-IR process analysis is a rapidly expanding, vigorous area within process analytical chemistry. Technological changes as well as various industry initiatives drive the changes in contact, noncontact, and liquid analysis, as well as the associated chemometrics, process communications, and process control [66].

REFERENCES

1. M. Tribus, *Deming's Definition of Management*, MIT Center for Advanced Engineering Studies, 1987.
2. W. Shewhart, *Economic Control of Quality in Manufactured Product*, Van Nostrand, New York, 1931.
3. V. C. Utterback, *Chem. Eng.*, June: 141–144 (1976).
4. J. Jutila, *In Tech*, July: 38–44 (1979).
5. S. M. Jacobs and S. M. Mehta, *Am. Lab.*, December: 15–22 (1987).
6. <http://www.fda.gov/cder/guidance/6419fnl.pdf>
7. J. B. Callis, D. L. Illman, and B. R. Kowalski, *Anal. Chem.*, **59**: 624A–635A (1987).
8. D. E. Honigs, *Am. Lab.*, December: 49–51 (1987).
9. W. E. Deming, *Mech. Eng.*, **66**: 173–177 (1944).
10. G. Kateman and F. W. Pijpers, *Quality Control in Analytical Chemistry*, Vol. 60, Chem. Anal. Ser. Monogr., P. J. Elving, J. E. Winfordner, and I. M. Kolthoff (eds.), John Wiley & Sons, New York, 1981.
11. D. M. Himmelblau, *Process Analysis by Statistical Methods*, John Wiley & Sons, New York, 1970.
12. K. J. Clevett, *Process Analyzer Technology*, John Wiley & Sons, New York, 1986.

13. K. Bakeev, *Process Analytical Technology*, Blackwell Publ. Co., Oxford UK (2005).
14. P. E. Mix, *The Design and Application of Process Analyzer Systems*, Wiley-Interscience, New York, 1984.
15. J. R. P. Clark, *Anal. Chem. Acta*, **190**: 1–11 (1986).
16. T. Hirschfeld, J. B. Callis, and B. R. Kowalski, *Science*, **226**: 312–318 (1984).
17. P. J. Cooper, *Cereal Food World*, **28**: 241–245 (1983).
18. H. Willard, L. Merritt, J. Dean, and F. Settle, *Instrumental Methods of Analysis*, 6th ed., Wadsworth, Belmont, CA, 1981.
19. D. E. Honigs, G. M. Hieftje, and T. Hirschfeld, *Appl. Spectrosc.*, **39**: 253–256 (1985).
20. B. G. Liptak and K. Venczel, *Process Measurement, Instrument Engineers' Hand-book*, rev. ed., Chilton, Radnor, PA, 1982.
21. D. C. Cornish, G. Jepson, and M. J. Smurthwaite, *Sampling Systems for Process Analyzers*, Butterworths, London, 1981.
22. T. Hirschfeld, D. Honigs, and G. Hieftje, *Appl. Spectrosc.*, **39**: 430–433 (1985).
23. G. J. Kemeny and D. L. Wetzel, Moisture: Study of a Lively Near-Infrared Diffuse Reflectance Spectrum, Paper 117, FACSS 13th Annual Meeting, September 28 to October 3, 1986, St. Louis, MO.
24. G. J. Kemeny and D. L. Wetzel, Differences in the Spectrum of Walter, AACC Annual Meeting, November 1–6, 1987, Nashville, TN.
25. J. Birth, *Appl. Spectrosc.*, **36**: 336 (1982).
26. K. G. Carr-Brion and J. R. P. Clarke, *Sampling Systems for Process Analyzers*, 2nd ed., Butterworth-Heinemann, Oxford, England, 1996.
27. R. E. Sherman, *Process Analyzer Sample Conditioning System Technology*, Wiley, New York, 2001.
28. R. E. Sherman, PAT, NESSI, and Sample Conditioning Systems: State of the Technologies and Future Directions, *Spectroscopy*, January: 11–13 (2005).
29. C. Jones and J. E. Reed, *Chem. Eng.*, **85**: 111–114 (1978).
30. C. Jones, *In Tech*, August: 51–54 (1982).
31. I. Landa, Advances in On-line NIR Instrumentation, Eastern Analytical Symp., October 1986, New York.
32. <http://www.brukeroptics.com/ft-nir/matrix-e.html>
33. G. J. Kemeny and D. L. Wetzel, Novel Very High Speed Near-Infrared Spectrometer System, Paper 175, FACSS 13th Annual Meeting, September 28 to October 3, 1986, St. Louis, MO.
34. G. J. Kemeny and D. L. Wetzel, Acousto-Optic Tunable Filter Based Analyzer System, Third International Diffuse Reflectance Conference, August 17–22, 1986, Chambersburg, PA.
35. E. W. Cziurczak, *Spectroscopy*, **5**: 10–11 (1990).
36. G. J. Kemeny et al., US Patent 5,039,855 (1991).
37. T. Hirschfeld, *Appl. Spectrosc.*, **39**: 1085–1086 (1985).
38. W. M. Wendlandt and H. G. Hecht, *Reflectance Spectroscopy*, Interscience, New York, 1966, p. 46.
39. G. Kortum, *Reflectance Spectroscopy*, Springer-Verlag, New York, 1969.
40. G. J. Kemeny and D. L. Wetzel, Unpublished results, Kansas State University, Manhattan, 1985–1986.
41. N. N. Die, *Mühle Mischfuttermitteltechnik*, **120**: 229 (1982).
42. H. Bolling and Z. Zwingelberg, *Mühle Mischfuttermitteltechnik*, **119**: 550–554 (1982).
43. D. L. Wetzel, On-Line Monitoring for the Control of Process, NIRA Symposium, 1984, Tarrytown, NY.
44. D. L. Wetzel, J. A. Levin, and E. S. Posner, Remote On-line Monitoring: Hardware, Software and Economics, NIRA Symposium, 1985, Tarrytown, NY.
45. G. J. Kemeny and R. Rubinovitz, Spectral Library for the Identification of Inorganic Materials Used in the Pharmaceutical Industry, PAT Conference, London, July 2004.
46. M. Blanco and M. A. Romero, Near-infrared Libraries in the Pharmaceutical Industry, *Analyst*, **126**: 2212–2217 (2001).
47. Y.-A. Woo, H.-J. Kim, and J. Cho, *Microchem. J.*, **63**: 61–70 (1999).
48. A. S. El-Hagrasy, H. R. Morris, F. D'Amico, R. A. Lodder, and J. K. Drennen 3rd, *J. Pharm. Sci.*, September **90**: 1298–307 (2001).
49. R. D. Maesschalck, F. C. Sanchez, D. L. Massart, P. Doherty, and P. Hailey, *Appl. Spectrosc.*, **52**: 725–731 (1998).
50. G. J. Kemeny, Proposed Standardization of Near-Infrared Blend Monitoring, International Diffuse Reflectance Conference, Chambersburg, 2004.

51. R. A. Crocombe, MEMS Spectroscopy Moves Process Spectroscopy into a New Dimension, *Spectrosc. Eur.*, **16**: 16–19 (2004).
52. S. Watano, H. Takashima, Y. Sato, K. Miyunami, and T. Yasutomo, IR Absorption Characteristics of an IR Moisture Sensor and Mechanism of Water Transfer in Fluidized Bed Granulation, *Adv. Powder Technol.*, **7**: 279–289 (1996).
53. J. Rantanen et al., *AAPS Pharmsci. Tech.*, **2**: (2001): article 10.
54. J. Rantanen, Near-Infrared Reflectance Spectroscopy in the Measurement of Water as a Part of the Multivariate Process Monitoring of Fluidized Bed Granulation Process, Academic Dissertation, Helsinki, 2000.
55. M. Derksen, P. Van de Oetelaar, and F. Maris, The Use of Near-Infrared Spectroscopy in the Efficient Prediction of the Residual Moisture of the Freeze-Dried Product, *J. Pharm. Biomed. Anal.*, **17**: 473–480 (1998).
56. I. Last and K. Pebble, Suitability of Near-Infrared Methods for the Determination of Moisture in a Freeze-Dried Injection Product Containing Different Amounts of the Active Ingredient, *J. Pharm. Biomed. Anal.*, **11**: 1071–1076 (1993).
57. W. B. Davenport and W. L. Root, *Random Signals and Noise*, McGraw-Hill, New York, 1958.
58. G. J. Kemeny, R. Rachlis, H. Mark, and J. Workman, Effect of Sample Motion in NIR Analysis, Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, February 22–26, 1988, New Orleans.
59. P. J. Mann, *Food Eng.*, July: 130–132 (1977).
60. S. Kerry, *Control Instrument.*, **13**: 41–42 (1981).
61. C. T. Andersen, R. E. Rice, and R. C. Polta, *Evaluation of Noncontact Moisture Analyzers*, Metropolitan Waste Control Commission, St. Paul, MN, 1979–1980.
62. <http://www.ndcinfrared.com>
63. J. A. Sneller, *Mod. Plastics*, **30**: 34–37 (1987).
64. K. I. Hildrum, B. N. Nilsen, F. Westad, and N. M. Wahlgren, In-Line Analysis of Ground Beef Using a Diode Array Near Infrared Instrument on a Conveyor Belt, *J. Near Infrared Spectrosc.*, **12**: 367–376 (2004).
65. <http://www.brukeroptics.com/ft-nir/matrix-e.html>
66. J. Workman, M. Koch, and D. Veltkamp, Process Analytical Chemistry, *Anal. Chem.*, **77**: 3789–3806 (2005).

38 Detection of Counterfeit Currency and Turquoise

Donald A. Burns

CONTENTS

38.1	Introduction	761
38.2	Currency.....	761
38.2.1	Background	761
38.2.2	Experimentation	762
38.2.3	Identification of Best Wavelengths (Unconventional Approach).....	762
38.2.3.1	Challenge #1	763
38.2.3.2	Challenge #2	763
38.2.3.3	Challenge #3	764
38.2.3.4	Challenge #4	766
38.3	Turquoise	768
38.3.1	Background	768
38.3.2	Experimentation	769
	Further Reading	772
	References	773

38.1 INTRODUCTION

A near-infrared (NIR) spectrum can reveal information that counterfeiters don’t anticipate. As such, it can be used to detect fakes in those materials where the additional spectral information above the visible region differs from that in genuine materials. This chapter addresses efforts to thwart those who would print counterfeit currency or offer colored plastic to tourists seeking to purchase turquoise jewelry. Some of this work was done by the author during his tenure as a staff member at Los Alamos National Laboratory (LANL).

38.2 CURRENCY

38.2.1 BACKGROUND

When a researcher at LANL described his efforts to embed a microdot (containing information confirming its legitimacy) in new currency, two questions arose: “What about existing currency?” and “Could NIR be a route to the detection of counterfeit currency?” The technique had already been applied to paper [1–5] and a study by Dale and Klatt [6] had employed principal component analysis (PCA) to currency, albeit only new and aged samples.

A request to a local bank for samples of counterfeit currency was denied, a bank officer pointing out (a) the illegality of possessing any and (b) the requirement that they must immediately turn in to the Secret Service (SS) whatever they receive. Accordingly, a formal request was made to the nearest office of the SS, which forwarded it to their Forensic Services Division in Washington, DC.

Two months later, a registered letter arrived containing ten counterfeit \$20 notes, and an analytical plan was formulated.

38.2.2 EXPERIMENTATION

Prior to scanning the bills, two major concerns had to be addressed:

1. What is the best spectral range? Possibilities were 400–2500 (the entire Vis-NIR), 700–1100 (the NIR), or 1100–2500 (the conventional NIR). The 400–750 range was considered the least useful, because humans are likely to be able to see whatever differences occur here.
2. Should the entire bill be scanned, or only selected parts? A full scan would necessarily include both paper and ink, while an appropriate mask could limit the observation to areas containing no ink (to answer the obvious question of whether paper, ink, or both would be required for discrimination). Also, if the entire bill was to be scanned, then some sort of sample moving mechanism would have to be designed and constructed.

Actual scanning was done with a Foss/NIRSystems Model 6500 spectrophotometer with a remote reflectance attachment at the end of a 5-ft fiber-optic cable. The reflectance module was turned upside down so its 2" × 2" window faced upward. The window was covered with a piece of white Teflon containing a rectangular opening 5 × 12 mm, thus forming a mask to define the scanned area. A block of white Teflon was used as backing, in other words, the currency was sandwiched between the two pieces of Teflon. For the "paper only" study with the \$20 bills, the area visible to the instrument was on the front of the bill near the left side, near the center (just to the right of the portrait), and near the right side. This positioning avoided having any inked area show through the mask.

Scanning was done over the entire visible/NIR range in 2-nm increments, but only the NIR portion was used for discrimination. Use of the vendor's software (IQ²) showed that 100% discrimination between genuine and counterfeit bills was obtainable on the paper only. However, one need not necessarily employ such a sophisticated software package. A case in point is a spreadsheet approach to divining wavelengths.

38.2.3 IDENTIFICATION OF BEST WAVELENGTHS (UNCONVENTIONAL APPROACH)

Seven genuine \$20 bills were scanned as described previously along with the ten bogus bills. Three scans of each bill provided 51 spectra. The standard plot of absorbance $\log(1/R)$ vs. wavelength is shown in Figure 38.1a. The data treatment consisted of taking the second derivative and subtracting the mean from each spectrum, thus emphasizing the differences between the two groups (Figure 38.1b). To further emphasize these differences and help identify those specific wavelengths where the greatest differences occur, the scale was expanded (Figure 38.1c). The two largest "loops" are labeled A and B. Absorbances at each of two wavelengths (1454 and 1492 nm) were then recorded for all 51 spectra.

To show this as a graph with two clusters of points, a Lotus 123 spreadsheet was constructed with three columns: (a) absorbances at 1454 nm for all 51 spectra (the X-values), (b) absorbances at 1492 nm for the genuine bills (the Y₁-values), and (c) absorbances at 1492 nm for the counterfeit bills (the Y₂-values). The result was a cluster diagram (Figure 38.1d) wherein the separation is estimated to be about 17 Mahalanobis units. The vendor's discriminant software (IQ²), using the same spectra but different wavelengths, calculated a better (larger) Mahalanobis distance (>40 U). In either case, it is readily apparent that separation of the two clusters was complete.

When this study was presented at a seminar attended by government and banking officials, a number of challenges were issued. We'll let these challenges, and their individual resolutions, tell the rest of the story.

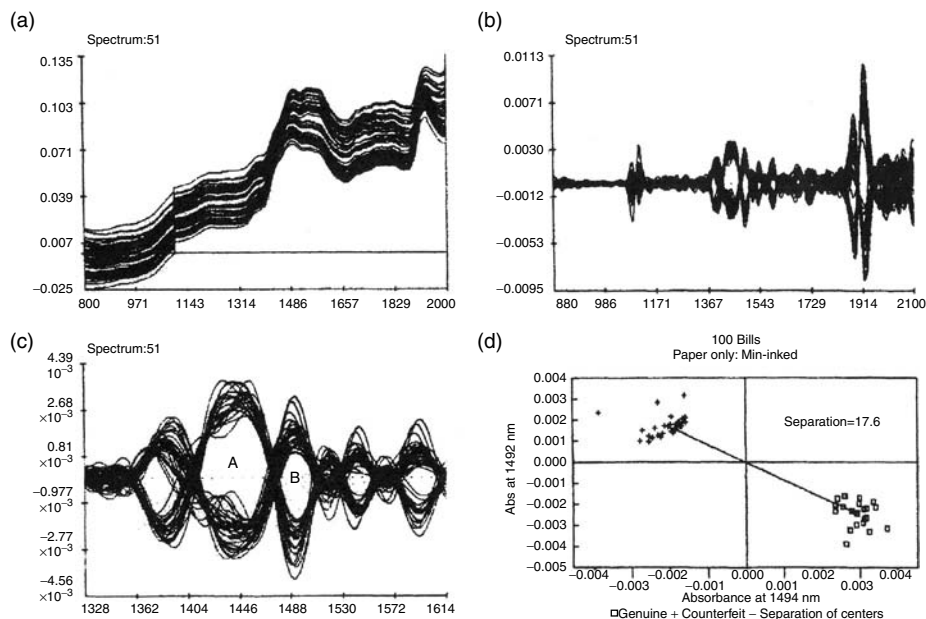


FIGURE 38.1 \$20 bills: (a) $\log(1/R)$ tracings of both genuine and counterfeit bills; (b) second derivatives with mean subtracted; (c) scale expanded to show major “loops”; (d) cluster diagram from loops A and B.

38.2.3.1 Challenge #1

It was postulated that the technique probably wouldn’t work on the recently discovered Supernote — a counterfeit \$100 bill that had been missed by existing detection devices at all 12 federal reserve banks [7]. Obtaining samples of these counterfeit \$100 bills wasn’t easy, but after several weeks of correspondence, 24 bogus bills became available at the local SS office.

It was decided to scan through the same mask (5×12 mm) that had been used with the counterfeit \$20 bills and look at an inked area on the back of the \$100 bill. Granted, there are many areas on the bills worthy of examination, and we did in fact evaluate such inked areas as (a) the black seal on the front, left of center, (b) the green seal on the front, right of center, (c) the portrait on the front, and (d) the picture of the tower on the back. After due consideration, we chose to zero in on the tower.

First, tracings in $\log(1/R)$ mode of the 24 counterfeit bills were made (Figure 38.2a) followed by tracings of 30 genuine \$100 bills kindly lent by a local bank (Figure 38.2b). The 30 genuine bills consisted of a random mix without regard to date of issue. The two groups were then displayed together (Figure 38.3a) and major differences emerged in the lower third of the family of tracings. The greatest difference occurred at 900 nm and this is emphasized in the first derivative tracing (Figure 38.3b).

It was incorrectly assumed that the counterfeit \$100 bills were a mix of Supernotes and other bogus bills. But an expansion of the scale of the second derivative revealed the presence of two groups (Figure 38.4). Suspecting that we had been given a mixture of Supernotes and other (non-Supernote) counterfeit bills, a query to the Washington office revealed that they were *all* Supernotes, and that we had identified two different ink recipes used by the counterfeiters.

38.2.3.2 Challenge #2

Our limited database (30 genuine bills) was regarded as insufficient, and we were also questioned as to the age of the bills. Another 30 genuine bills were borrowed from the bank and this time we kept track of the date each was printed. Included were bills from the 1960s to 1990s. Granted, age made a difference. However, even though some of the oldest bills had a spectrum similar to the counterfeit bills in the $\log(1/R)$ mode, their descending peak in the first derivative mode appeared

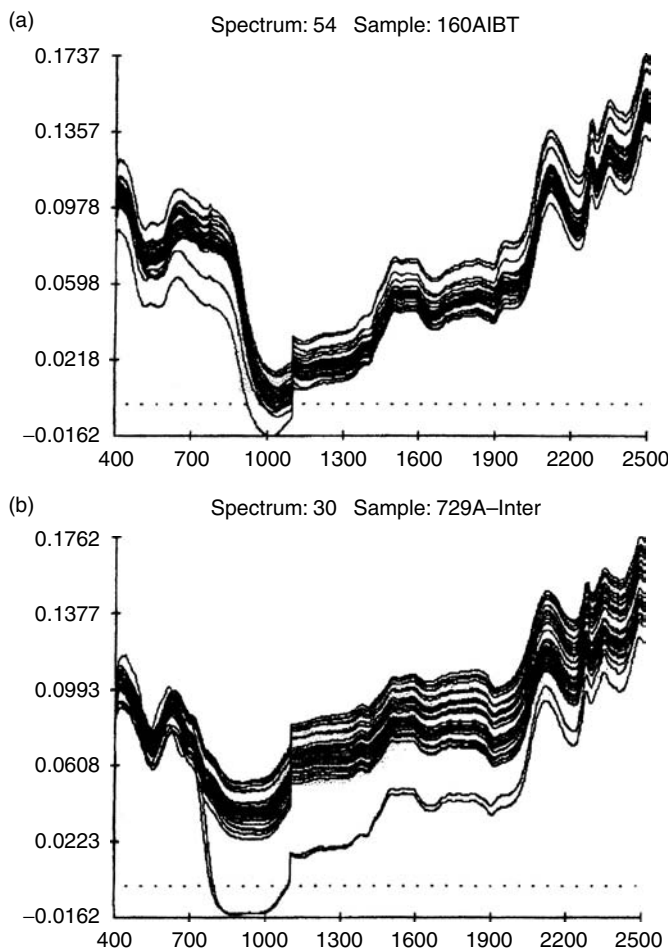


FIGURE 38.2 \$100 bills: (a) $\log(1/R)$ tracings of 24 “Supernotes”; (b) $\log(1/R)$ tracings of 30 genuine bills (10 each old, new, and intermediate).

at a lower wavelength (recall Figure 38.2b) and they were easily distinguished from the counterfeit bills (Figure 38.5). Moreover, when all 84 bills (24 counterfeit plus two groups each of 30 genuine) were displayed together (Figure 38.6), it was apparent that, old or new, the genuine bills could be easily distinguished from the Supernotes. Bills printed in the 1980s and 1990s were tightly clustered (Figure 38.7) while bills printed in the 1960s and 1970s formed another cluster, well separated from the counterfeit money and reasonably separated from the more recently printed genuine bills.

Many different approaches exist in the legal printing of currency to permit detection when inept counterfeiters are unaware of their presence [8]. But it is a cat-and-mouse game of production and deception that continues indefinitely: as each deterrent is devised, the counterfeiters beat it. The rest of this section confirms what is well known by government officials. Confidentiality precludes the naming of sources of additional counterfeit \$100 bills; they are identified only as “source A” and “source B.”

38.2.3.3 Challenge #3

“Source A” tested our algorithm with a special \$100 bill that closely matched the genuine group but was in fact counterfeit (Figure 38.8a). Although a major difference occurred near the top of the

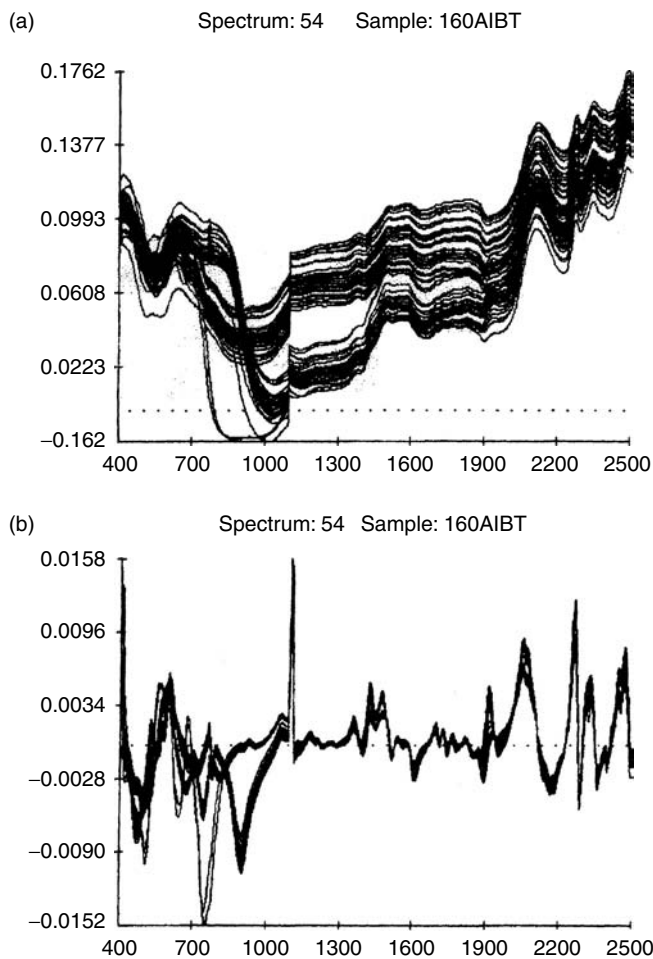


FIGURE 38.3 \$100 bills, composite of 24 counterfeit and 30 genuine: (a) $\log(1/R)$ tracings; (b) first derivatives.

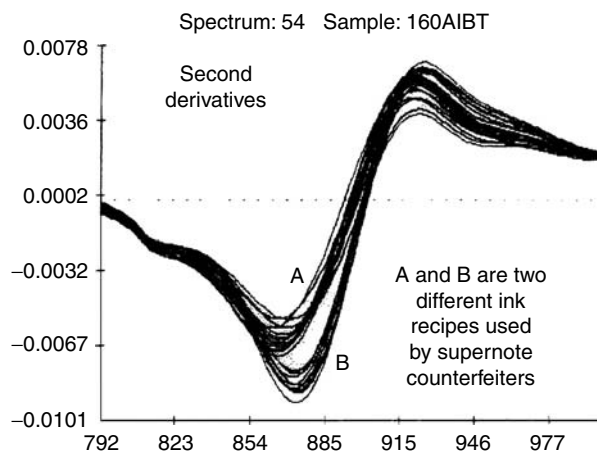


FIGURE 38.4 \$100 bills: second derivatives of 24 Supernotes depicting two different ink recipes.

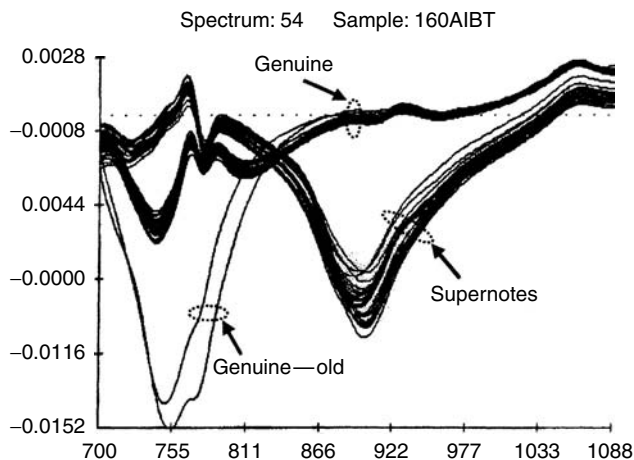


FIGURE 38.5 \$100 bills: first derivatives of mixture (24 Supernotes plus 30 genuine) showing effect of age.

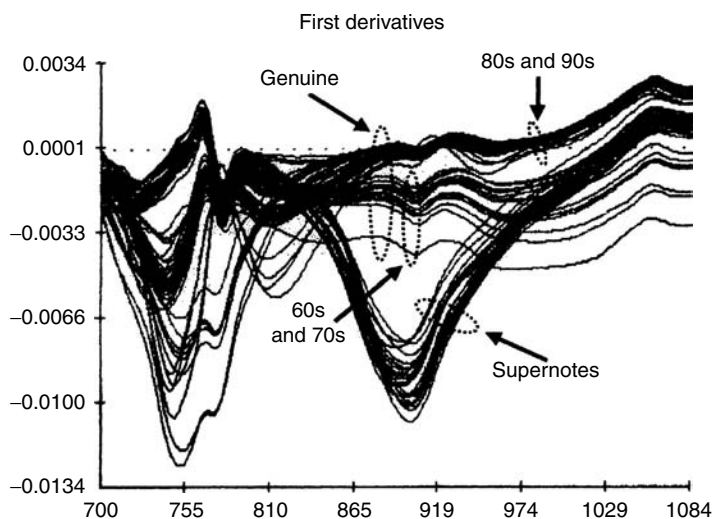


FIGURE 38.6 \$100 bills: first derivatives of larger mixture (24 Supernotes plus 60 genuine) showing effect of age.

visible region, it was not apparent, even to the trained eye. The second derivative tracing in this region, however (Figure 38.8b), shows how easy it was to identify. Likewise, this bill had many of the characteristics of the Supernotes (Figure 38.8c) but its second derivative tracing (Figure 38.8d) provided a means of distinguishing it.

38.2.3.4 Challenge #4

Finally, “source B” presented us with 24 counterfeit \$100 bills, and on the first run two were incorrectly identified as genuine. Now it was time to do some fine-tuning. Figure 38.9 details the sequence of events that led to a modification of the algorithm to account for these new found differences. In the initial $\log(1/R)$ tracing (Figure 38.9a), these superb counterfeit bills looked much like a typical group of 30 genuine bills, except for the offset (not very useful in distinguishing them). Even the first derivative (Figure 38.9b) did not help much, because a couple of the genuine bills showed a similar peak (albeit smaller) in the area of greatest difference. But the second derivative

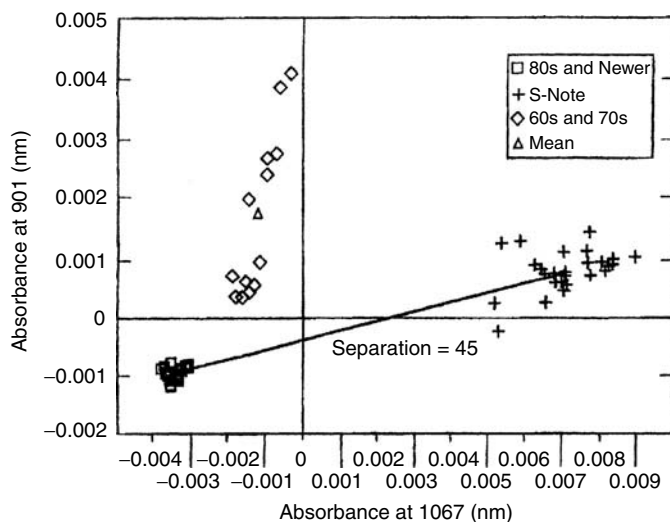


FIGURE 38.7 \$100 bills: cluster diagram showing excellent separation of Supernotes from different groups of genuine currency.

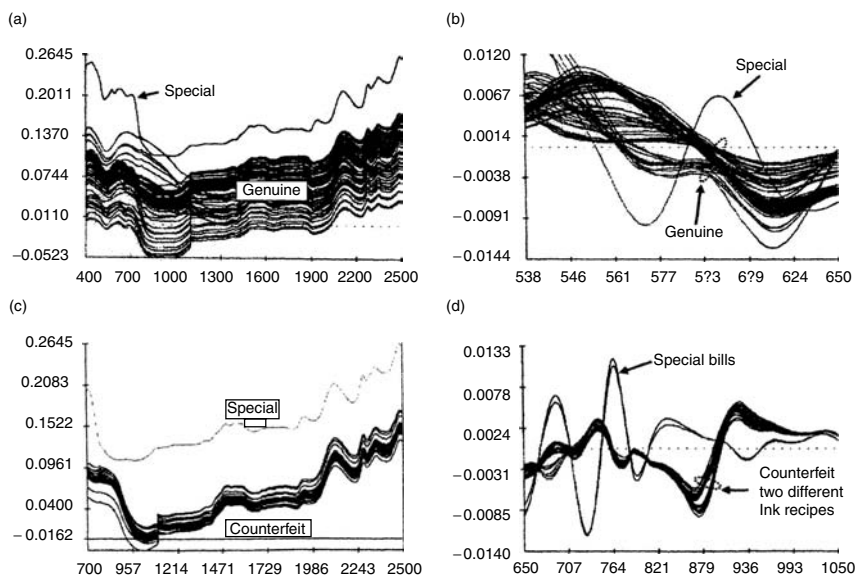


FIGURE 38.8 Comparison of special counterfeit \$100 bill with 60 genuine bills and 24 Supernotes: (a) $\log(1/R)$ tracings; (b) second derivatives; (c) $\log(1/R)$ tracings; (d) second derivatives.

(Figure 38.9c) clearly separated these excellent counterfeits from the genuine bills, whether old or not. A difference also occurred in the high end of the visible region (note peaks A and B), but their proximity was too close for reliable distinction.

Much more could be (and has been) written on this subject. Suffice it to say here that NIR is a proven route to the detection of counterfeit currency, and because of the relatively and unavoidably small database, work should be ongoing to thwart the efforts of counterfeiters. LANL offered this work as an R & D 100 Contest entry in 1996 under the title “SuperScan Counterfeit Currency Detector” but it was not among the top 100. Moreover, the FBI and SS seemed disinterested, saying *there is no problem*. However, an invention disclosure was filed and a patent [9] issued.

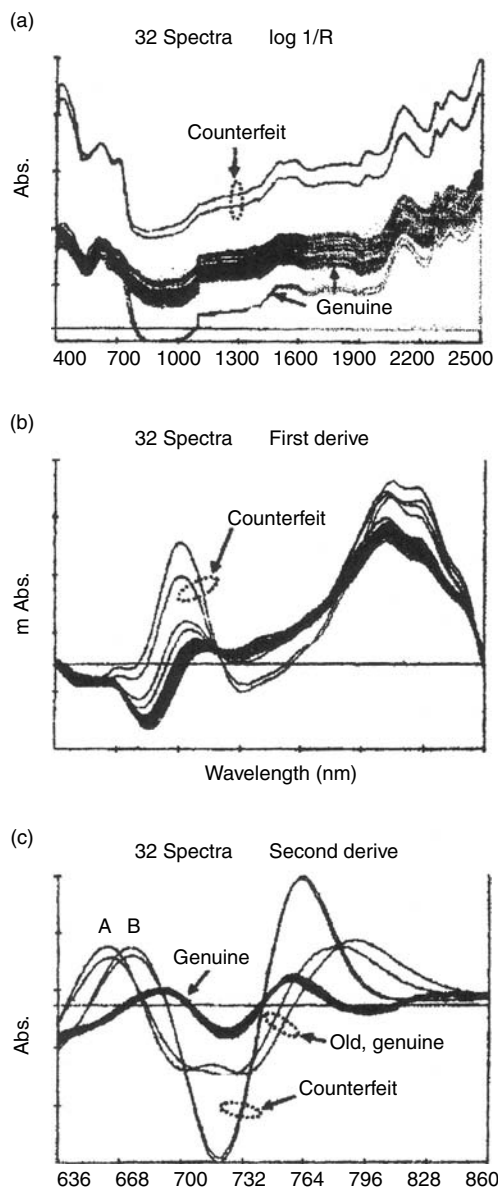


FIGURE 38.9 Comparison of two excellent counterfeit \$100 bills with several genuine bills: (a) $\log(1/R)$ tracings; (b) first derivatives; (c) second derivatives.

38.3 TURQUOISE

38.3.1 BACKGROUND

It all started with a story on the evening TV news. A method was being described for detecting fake turquoise, because (a) jewelry is big business in the Southwest — Santa Fe, Albuquerque, and Taos in New Mexico and (b) the tourists ought to be protected from being sold counterfeit versions of it. The TV screen showed a blue material sold as turquoise, but suspected of being colored plastic, undergoing a “test” in a crucible suspended over a torch. If the material turned black, it was not turquoise; if it did not, then it might have been the genuine material. Shortly after this broadcast, articles appeared in local newspapers [10,11] calling readers’ attention to the dishonesty of some vendors.

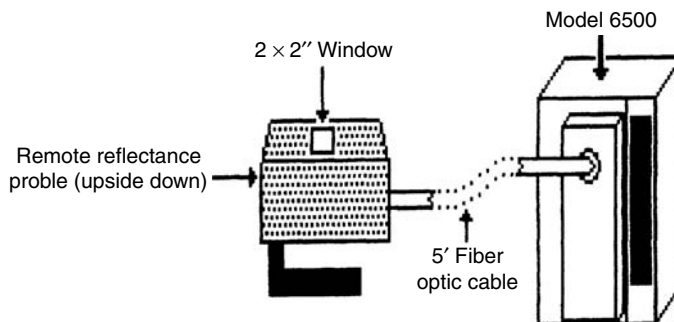


FIGURE 38.10 Diagram of instrumentation for turquoise analysis.

The mention of the word *plastic* caught the author's attention. It is well known that NIR has been used successfully for years to identify various plastics in a totally nondestructive manner. Why, then, not use it to differentiate plastic turquoise lookalikes from the real McCoy? Accordingly, samples were solicited from local jewelry stores and jewelry supply houses. This wasn't any easy task. More than once I was immediately shown the front door when I asked for samples of fake turquoise. (The nerve of even suggesting that a jeweler might offer other than the genuine material!) But one sympathetic dealer and a local consultant/collector eventually lent me a few stones.

38.3.2 EXPERIMENTATION

The same instrumentation used in the counterfeit currency study was used here (Figure 38.10). The remote reflectance probe of the FOSS/NIRSystems Model 6500 was turned upside down so samples could be placed on its 2" x 2" window, always covered with a black cloth during scanning.

The first few scans were quite encouraging. Tracings of the three plastic fakes were significantly different from both natural turquoise and some treated versions (Figure 38.11). There are many different treatments used to embellish turquoise stones: polished, fracture-scaled, enhanced, stabilized, and so forth. The Grolier encyclopedia [12] has pointed out that both now and in the past centuries far more substitute and artificial gemstones have been in circulation than natural, properly identified stones. And this surely applies to turquoise. It is quite legal to sell treated turquoise stones as long as they're not identified as *genuine* turquoise.

Derivative treatments generally identify fakes easily (Figure 38.12). These second derivatives reveal distinct peaks for plastic substitutes in a region where natural turquoise as well as many treated versions are relatively flat. Employing NIRSystems' IQ² software, the "Match by Distance" of 12 samples of natural, treated, and fake stones had average Mahalanobis distances within each group of 1.26, 1.05, and 0.77, respectively. On another group of 14 samples, all correctly identified by distance, these numbers were 1.06, 1.44, and 0.75, respectively. Match by correlation was always less accurate.

Cluster diagrams generally provide a good visual presentation of group separations. Figure 38.13 reveals a separation of 14 Mahalanobis distances between plastic fakes and natural stones, using wavelengths identified in a 2-wavelength discriminant analysis program followed by LOTUS 123 graphics. Figure 38.14 depicts the best two of a 3-wavelength discriminant analysis using Technicon's (13) IDAS software. Results are shown in the following table:

Group	Name	Mahalanobis	Distances from
1	Fake	1-2	19.2
2	Stabilized	1-3	13.4
3	Natural	2-3	29.3

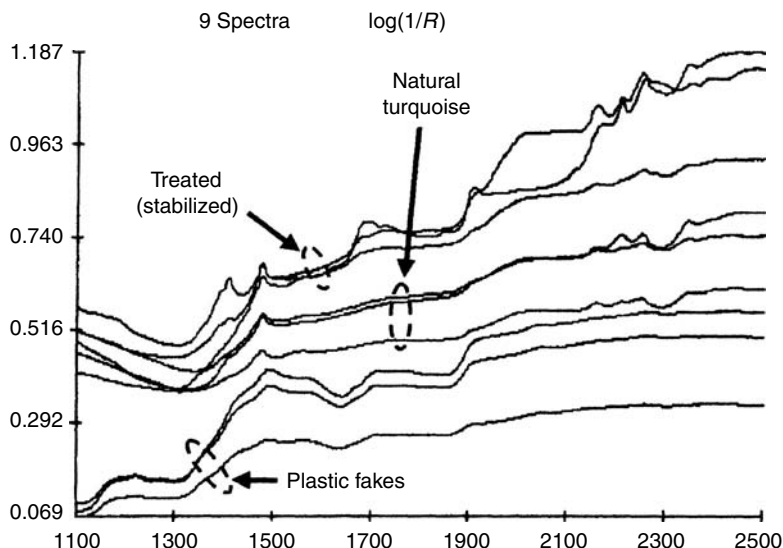


FIGURE 38.11 Tracings in $\log(1/R)$ mode of first nine samples.

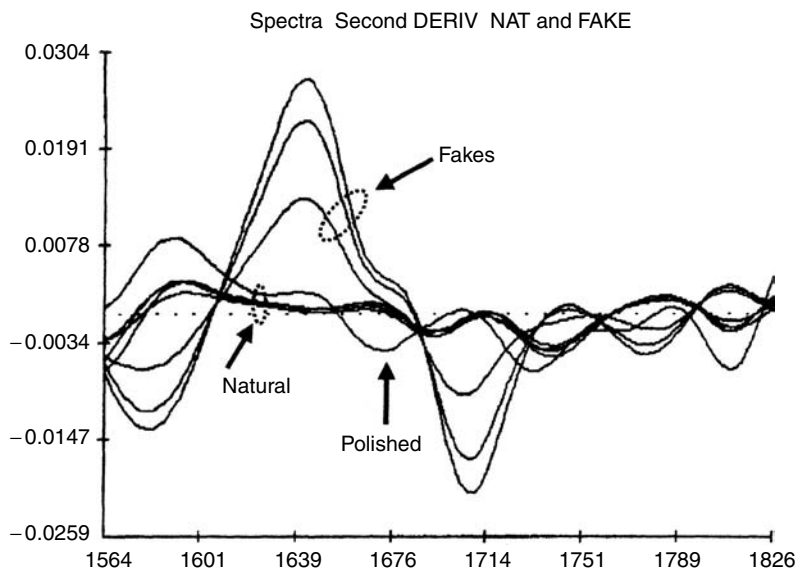


FIGURE 38.12 Second derivative tracings of natural and fake turquoise; one sample was polished.

Turquoise is regarded as the mineral cuprous aluminum phosphate with the following formula:



Because the Cu and the Al show no appreciable structure in the NIR region, it is likely that we are looking at P—O or P=O bonds along with the usual —OH group. But there is a close cousin of turquoise known as variscite whose formula is



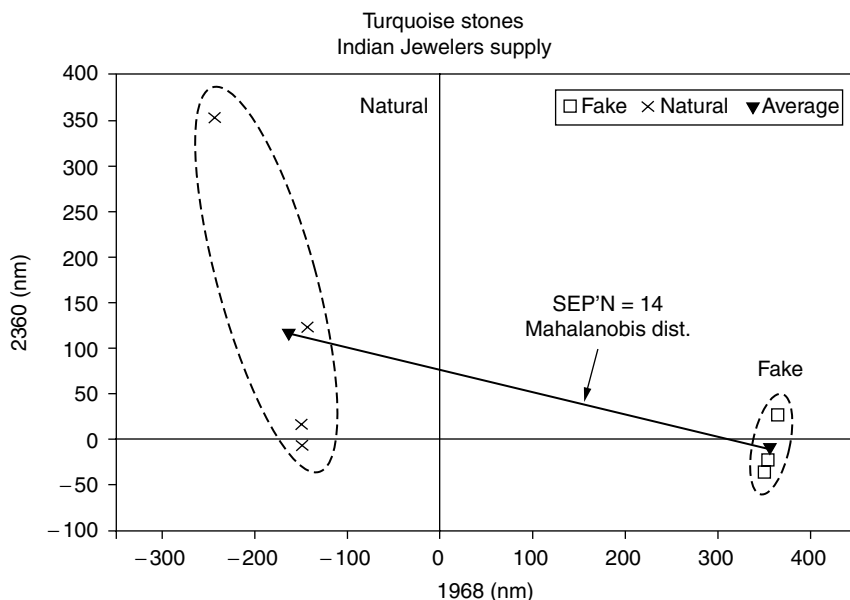


FIGURE 38.13 Cluster diagram showing excellent separation of natural and fake turquoise.

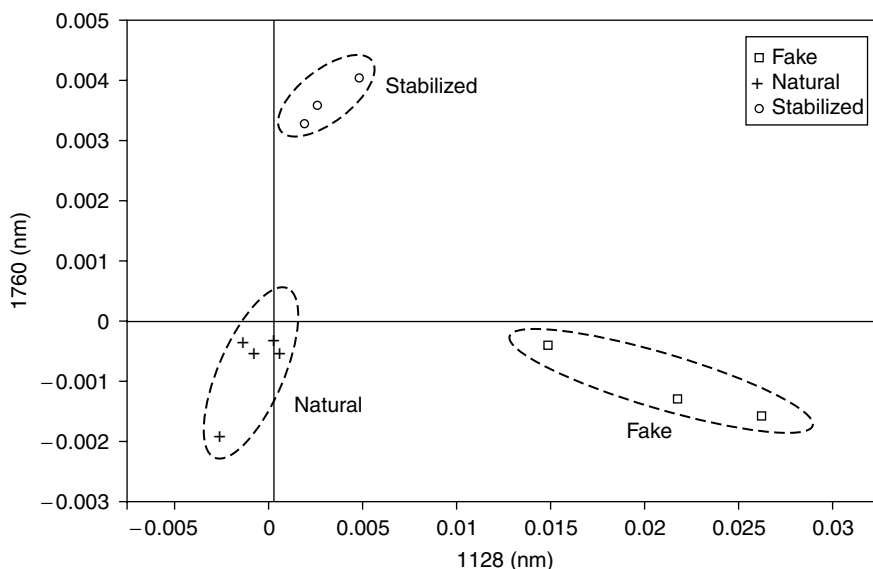


FIGURE 38.14 Cluster diagram from best two of three-wavelength discriminant analysis showing separation of three classes of samples.

A New Mexico jewelry wholesaler wondered if this material, sometimes sold as genuine turquoise, could be distinguished from true turquoise. Figure 38.15 shows that it can usually be distinguished from both natural turquoise and plastic fakes.

Finally, unlike many materials that can be altered in form to accommodate an analytical system or procedure, gems are confined to the realm of nondestructive testing. An article by Armstrong, Wang, Beesley, and Ribinovit [14] deals with turquoise and rubies in this connection. These workers

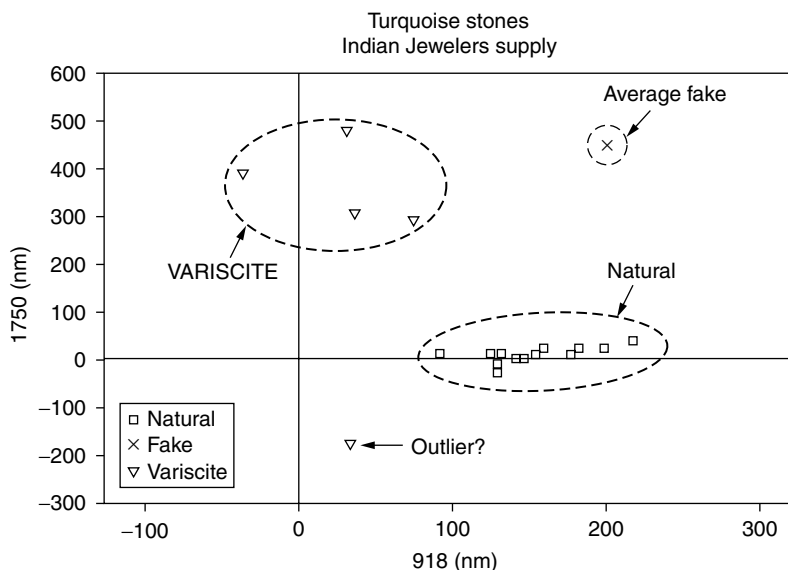


FIGURE 38.15 Cluster diagram resolving the question “Can VARISCITE be distinguished from natural turquoise and counterfeit stones?”

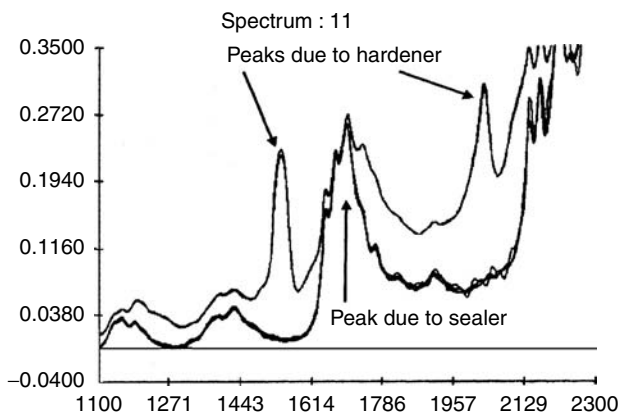


FIGURE 38.16 Easy detection of adulteration of turquoise by epoxy resins for so-called “fracture-sealed” treatment.

pointed out that the current analytical challenge is the detection of organic fillers used to alter the appearance of surface-reaching fissures. These fillers are typically epoxy resins, and Figure 38.16 reveals how easy they are to detect. Manufacturers of turquoise jewelry very often use this kind of treatment to enhance their product.

In this chapter we have attempted to show how NIR can be applied to the field of counterfeiting. The examples of currency and turquoise gemstones are intended to whet the appetite of those who see a wide-open field here and are willing to further explore it.

FURTHER READING

1. M. N. Adams, Conspiracy Against the Dollar, *Reader's Digest*, 69, March, 1995.
2. Z. Pamela, High-tech ink to be added to U.S. bills, *Chemical & Engineering News*, October 9, 1995, pp. 7–8.

3. Checking Currency Paper on the Fly, *Lasers & Optonics*, December 1995, pp. 29–30.
4. The Science of Money, *Discover*, October 1998.

REFERENCES

1. C. M. Paralusz, ACS National Meeting, Abstr #85, 1987, Denver, CO.
2. F. A. DeThomas, Rocky Mt. Conference, Abstr #206, 1988.
3. C. M. Paralusz, *Appl. Spectr.*, 43: 1273 (1989).
4. P. J. Brimmer, Pittsburgh Conference, Abstr #1373, 1990, New York.
5. S. Monfre, Rocky Mt. Conference, Abstr #211, 1990.
6. M. D. John and N. K. Keon, *Appl. Spectrosc.*, 43: 1399 (1989).
7. F. Dannen and I. Silverman, The Supernote, *The New Yorker*, October 23, 1995, pp. 49–55.
8. R. Lipkin, New Greenbacks, *Science News*, 149: 58 (1996).
9. D. A. Burns, Detection of Counterfeit Currency, Patent Number: 5,757,001, May 26, 1998.
10. Fake jewelry comes under fire, *The Santa Fe New Mexican*, August 13, 1998.
11. Retired LA scientist develops device that can easily detect fake turquoise, *Los Alamos Monitor*, September 11, 1998.
12. Grolier Electronic Publishing Inc., Copyright 1992.
13. Technicon is now Bran + Luebbe.
14. D. W. Armstrong, X. Wang, C. R. Beesley, and R. Ribinovitz, Analysis of fissure-filled turquoise, emeralds, and rubies by near-infrared spectroscopy, *Amer. Lab.*, 31: 41–47 (October 1999).

39 Counterfeit Cigars: Can Near-Infrared Detect Them?

Donald A. Burns

CONTENTS

39.1 Introduction	775
39.2 Experimental	775
39.3 Conclusions	778
39.4 Additional Reading	779
References	779

39.1 INTRODUCTION

More than 3,000,000 counterfeit cigars are sold in the United States every year. These fakes are said to be “Cuban” but they are not. This preliminary study addresses the feasibility of near-infrared (NIR) spectroscopy for identifying the fakes.

39.2 EXPERIMENTAL

Following a query on the Internet, 14 samples were provided by a curious and concerned dealer in tobacco products: 7 each for Cuban and non-Cuban cigars. All but one came with its identifying band. They were scanned over the range 400 to 2500 nm on a Foss/NIRSystems Model 6500 spectrophotometer. Samples were positioned on a remote reflectance attachment over a white Teflon mask to limit the scanned area to a $2 \times \frac{1}{2}$ in. rectangle. A minimum of four scans of each cigar was made by rotating the sample around its long axis in 90° increments. No obvious differences could be seen between the two groups, particularly below 1100 nm, so the first derivatives were computed as shown in Figure 39.1. The slight differences between the two groups can be seen in the boxed areas.

The vendor’s IQ² software revealed encouraging discrimination when used in the match by distance mode. Although two data points were $>3\sigma$, all IDs were correct (Table 39.1).

Absorbances at the two best wavelengths were used to show the Mahalanobis separation (5.2 U) in a discriminant plot (Figure 39.2). All but one of the non-Cuban cigars were reasonably well separated from the Cuban group when plotted with only two wavelengths (Figure 39.3).

Another discriminant analysis program from a third-party vendor (Mark Electronics, Suffern, NY) identified three wavelengths that produced even better separation when plotted as three-dimensional clusters (Figure 39.4). This suggests that three wavelengths may be necessary and sufficient for distinguishing between Cuban and non-Cuban cigars.

One cigar was sacrificed (peeled apart) to determine the depth to which the NIR beam penetrated, that is, how many layers (leaves) of tobacco were contributing to the spectrum of the cigar. This

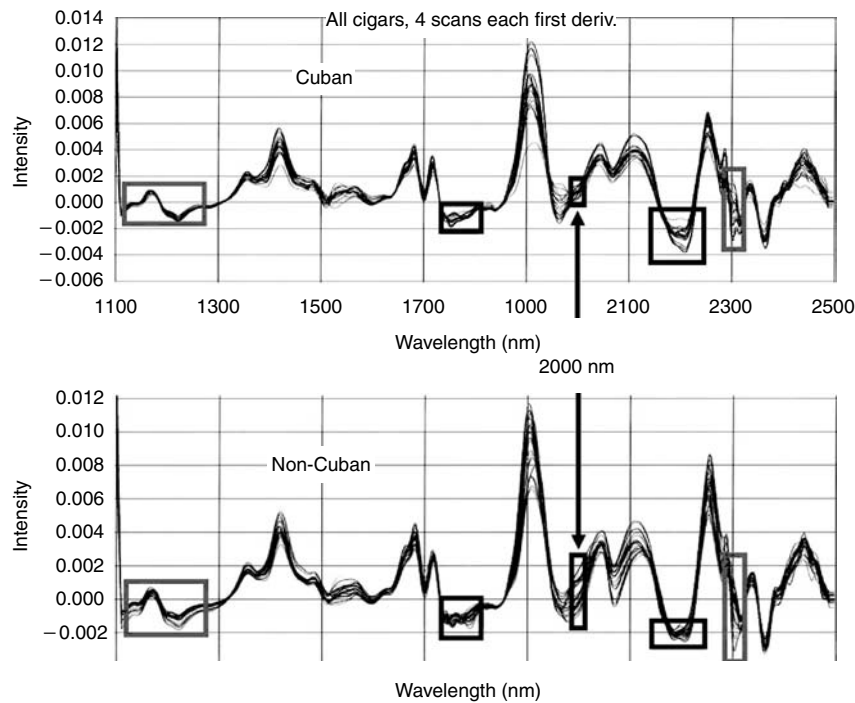


FIGURE 39.1 First derivatives of Cuban and non-Cuban cigars.

TABLE 39.1
Foss/NIRSystems Library Validation Report

Library:		CIGARS	
Number of products:		2	
Instrument:		6500	Math Treatment: 2nd Deriv
Scan Range:		400–2500	Selected: 1100–2466
Identify Correlation Threshold:		0.85	
Identify Distance Threshold:		6.00 σ	
Match by Correlation		Match by Distance	
		I	
		Y N	
Correct ID		II Y 56 0 56	
Y	41	N 2 0 2	
N	17	58 0 58	
	58	Type I: Sample ID Correct	
		Type II: Distance < 3.00 σ	

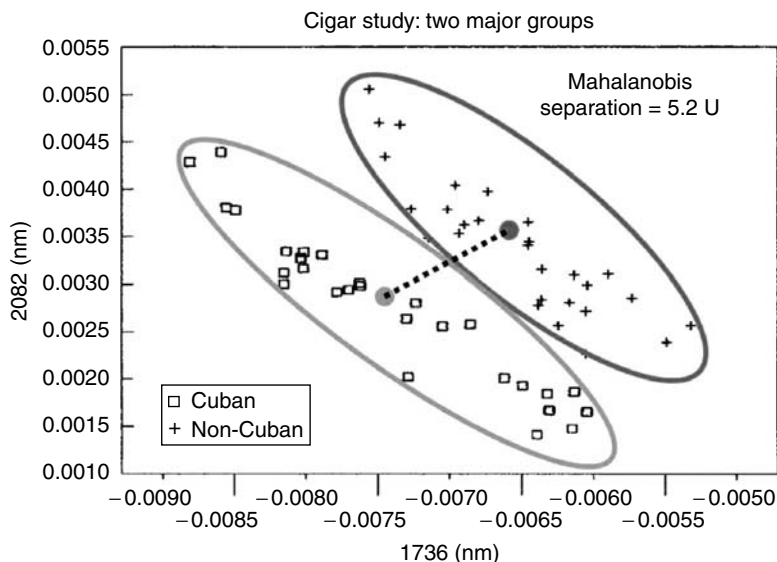


FIGURE 39.2 Cluster diagram of two major groups.

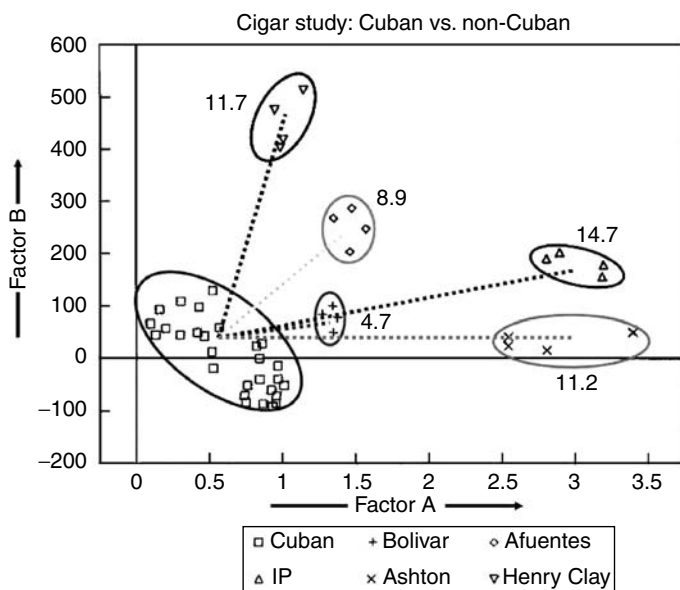


FIGURE 39.3 Cluster diagram — Cuban group vs. five non-Cubans.

could be estimated by observing the step-wise decreasing structure of the spectrum of a sheet of polyethylene plastic (Figure 39.5). Since the NIR beam descends through at least three layers, it seems unlikely that a counterfeiter could wrap a fake cigar with a single layer of Cuban tobacco and pass it off as a Cuban cigar.

When the “key” (identification) for the test samples (unknowns) was provided, none of them turned out to be among those that were used in the discriminant program. The identity of all 28 samples is given in Table 39.2.

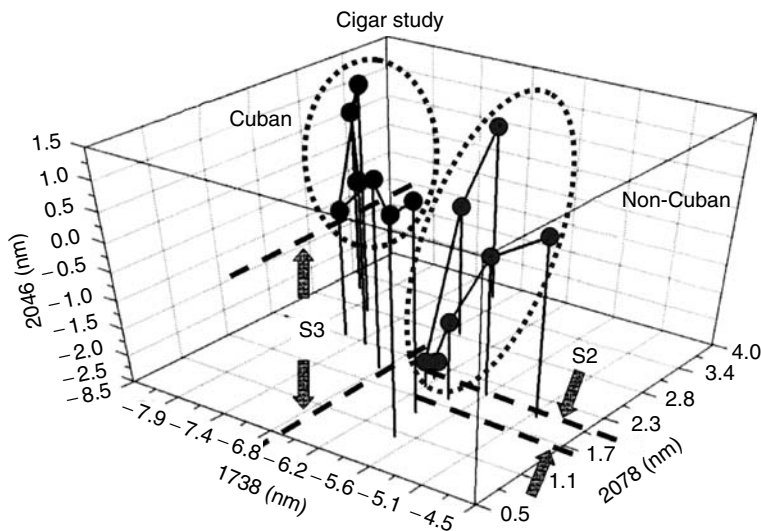


FIGURE 39.4 Three-dimensional cluster diagram; S2 = separation in two dimensions; S3 = separation in three dimensions.

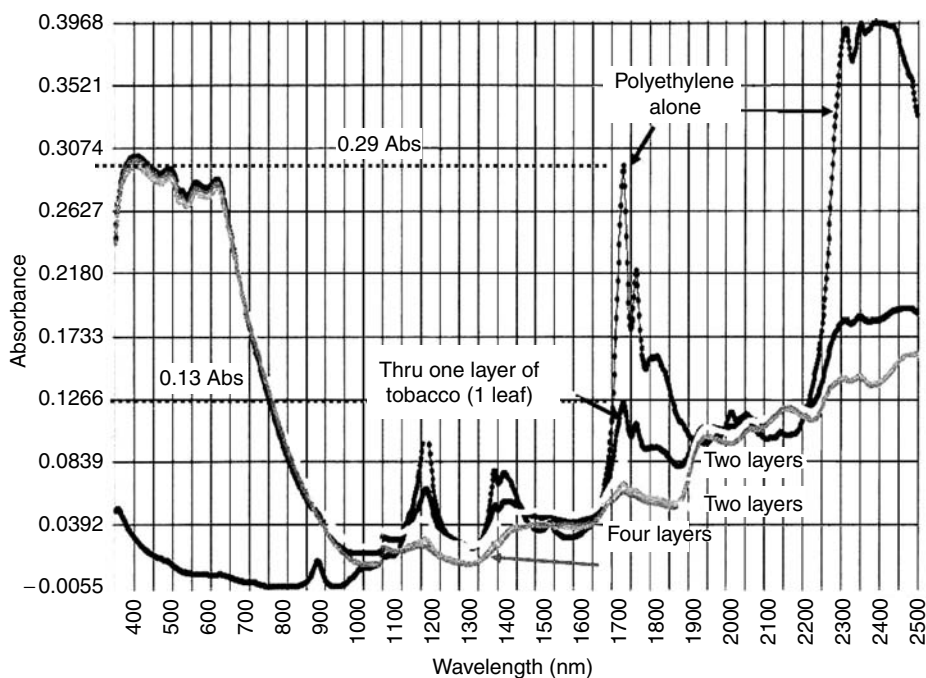


FIGURE 39.5 Penetration study.

39.3 CONCLUSIONS

Although this preliminary study strongly suggests that NIR can be a route to the detection of counterfeit cigars, it is apparent that a larger data set must be incorporated. A major difficulty with any study of counterfeit/fraudulent materials is finding a valid source of samples; those who possess such

TABLE 39.2
Identification of All Cigars: Top 14 Are Standards; Unknowns Are Lettered

	Cuban	Non-Cuban		
	Cohiba		Cifuentes	
	Unlabeled (short)		Punch	
	St Luis		Bolivar	
	Bolivar		Afuentes	
	Monte		IP	
	Unlabeled (long)		Ashton	
	Diplomat		Henry Clay	
		C	N	Origin
A	Habana Diplomatico #5	x		
B	Habana Hovo Epicure #1	x		
C	Padron Cafeteria		x	Nic
D	Habana SLR Serie A	x		
E	Trinidad Corona		x	Afr DR
F	FDA Maximo		x	Edu Hon DR
G	Vegas de Fonseca Churchill		x	Afr DR
H	Habana Cohiba Siglo 4	x		
J	Aristoff Perfecto		x	Ecu Nic DR
K	Moore & Bode Corona		x	USA
L	Habana Hoyo #2	x		
M	H Upmann Extra Finos		x	Ind DR
N	Habana Montecristi Robusto	x		
P	Bolivar Fuerte Cuban Corona		x	Ecu Hon Nic

C = Cuban; N = Non-Cuban; Afr = African; DR = Dominican Republic; Ecu = Ecuador; H = Honduras Ind = India; Nic = Nicaragua; USA = United States of America.

materials are usually reluctant to admit they are fakes. Equally important is knowing for certain that genuine samples are as labeled.

39.4 ADDITIONAL READING

There are a number of websites that deal with counterfeit cigars [1–5]. Most of them are essentially an offer to sell to subscribers, while some may address such items as the tax seal/stamp and various ways of packaging the cigars. None appears to deal with a scientific method of identification. Here, we take the first step with a spectrophotometric approach.

REFERENCES

1. "The Ultimate Counterfeit Cuban Cigar Primer," cigarnexus.com/counsel/counterfeit/index.html
2. "Mom's Cigars," davidoffcigars.com/genuinecounterfeitcuban.html
3. "Spotting a Fake Cuban Cigar," puroreycigars.com/Fakes.htm
4. "Franklin Liquors Cigar Page/Cuban Counterfeit Lab," franklinliquors.com/cubancigar.htm
5. "Counterfeit Gallery," cigaraficionado.com/Cigar/Aficionado/Counterfeit/countgal.htm

40 Local Methods and CARNAC-D

Tony Davies and Tom Fearn

CONTENTS

40.1	Introduction	781
40.2	Local Methods	782
40.2.1	CARNAC	782
40.2.2	Locally Weighted Regression (LWR)	782
40.2.3	LOCAL	783
40.2.4	Locally Biased Regression (LR)	783
40.3	CARNAC-D	783
40.3.1	Compression Methods	784
40.3.1.1	Fourier Transformation (FT)	784
40.3.1.2	Wavelet Transformation (WT)	784
40.3.2	Modification	785
40.3.2.1	Multiple Regression Method	785
40.3.2.2	Vector of Coefficients Method	786
40.3.3	Similarity	786
40.3.3.1	Linear Regression	786
40.3.3.2	Quadratic Regression	786
40.3.4	Outlier Tests	788
40.3.4.1	Quartile Method	788
40.3.4.2	Most Distant Samples Method	788
40.3.5	Determination of Weighted Mean	788
40.3.6	Optimization	789
40.3.7	Representative Results	789
40.3.7.1	Results from a Small Database	789
40.3.7.2	Results from a Large Database	789
40.3.8	Future Prospects for CARNAC-D	790
Appendix:	Carnac and CARNAC	792
40.A.1	The Long History of Carnac, Brittany, France	792
40.A.2	A Short History of CARNAC	792
References	794

40.1 INTRODUCTION

“Local methods” is a general term for a system for quantitative analysis that searches a database of samples with known spectral and analytical data for a small number of samples that are very similar in spectral terms to a sample for which only the spectrum is known. The analysis of the unknown sample is then inferred from the analysis of the similar samples. This is a very different approach

to the common methods (MLR, PCR, PLS, ANN, etc.) utilized in near-infrared (NIR) quantitative analysis, which use a model building “calibration” step prior to the analysis of unknowns.

The idea dates back at least to Jensen and Martens [1], who applied it to fluorescence data derived from cereals. Several different approaches have been described for application to NIR spectroscopy [2–6] and we will give a brief description of each, followed by a more detailed description of the method we have named “CARNAC-D.”

40.2 LOCAL METHODS

All local methods start with a database of analyzed (NIR spectrum and analyte values) samples and one or more samples with known NIR spectrum but unknown analyte values. The unknown samples will be compared individually with the database in order to infer their analyte values. Andersson, Osborne, and Wesley [2] have contributed a very useful discussion on the mathematical basis for local methods.

40.2.1 CARNAC

Comparison Analysis using Restructured Near-infrared And Constituent data (CARNAC), formulated in 1983 and published in 1988 [3], was the first application of local methods to NIR data. The basic idea was to search the database to find spectrally very similar samples compared to the sample with known spectrum but unknown analysis and predicted the analyte values as a weighted average of the selected sample. In order to make the process faster and in the hope of concentrating information, the original spectra were transformed to the Fourier domain; this allows a reduction from 700 data points in the original spectra to 100 pairs of Fourier coefficients. This technique was brought to NIR spectroscopy by Fred McClure (see Chapter 6) with whom this first version of CARNAC was developed. It is obvious that this simple idea will be dominated by the major analyte (or absorber) in the samples. To overcome this problem, in the setting up procedure of CARNAC, for each analyte under consideration, a set of coefficients are found that will emphasize that analyte. These are used to modify the database by multiplication of each point in the spectrum by the appropriate coefficient.

40.2.2 LOCALLY WEIGHTED REGRESSION (LWR)

Naes, Isaksson, and Kowalski [4] put forward the use of “locally weighted regression” in 1990. It is a local version of principal component regression. The spectra in the calibration database are transformed to scores on a chosen number of principal components, preferably after some pretreatment. Given an unknown, the same scores are computed from its spectrum, and the similar samples are selected as the closest ones in this principal component space. In computing this distance the scores may be weighted according to their correlation with the analyte [5]. A local calibration equation is then computed by multiple regression of the known analyte values on the scores for these similar samples, possibly weighting the samples according to their distance from the unknown. This equation is used to make a prediction for the unknown. In the original procedure, only the regression step of the principal component regression is local; the scores come from the global principal components analysis. A version in which the principal components are recomputed locally has subsequently been proposed [6]. Tuning the method involves optimizing the number of principal components, usually fewer than would be needed for a global calibration, and the number of similar samples in the local calibration. A more recent suggestion [6] is that, in LWR, similar samples may be selected by a criterion that combines squared distance in the principal component space and squared difference between analyte values. The analyte value for the unknown is initially estimated using a global calibration, though this may be followed by an iteration using local predictions. The relative weighting of the two parts of the distance measure is another parameter that may be optimized.

Sinnaeve, Dardenne, and Agneessens [7] published a study of global and local calibrations in which the local version used PCA distances to select similar samples but used PLS to produce a calibration from the selected samples. It thus shares features with both LWR and with LOCAL, described in the next section.

40.2.3 LOCAL

LOCAL (US Patent No. 5,798,526) [8], has a very important advantage for many users that it is implemented in commercially available software designed specifically for NIR calibration. If LWR is the local version of principal component regression, then LOCAL is the local version of PLS. Similar samples are selected using the correlation between the spectrum of the unknown and that of each database sample, as in CARNAC but using the actual spectra, possibly with some pretreatment, and reduction in the number of spectral data points, rather than a heavily modified version. The local calibration step is much more sophisticated for LOCAL than for the other two methods. A full PLS calibration is performed using the selected samples, and used to predict the unknown. The problem is that this requires a fully automatic method for deriving PLS calibrations, rapidly (and so preferably not using cross-validation to choose the number of factors) and with no user intervention. The solution adopted is quite a complex procedure employing a weighted average of several PLS calibrations. Several successful applications of LOCAL can be found in the literature [9–12].

40.2.4 LOCALLY BIASED REGRESSION (LR)

Locally biased regression [13] identifies samples similar to the unknown using distance in a two-dimensional space. One axis of this space is a prediction of the analyte using a global PLS calibration. The other is a score on a factor designed to capture spectral variation not associated with the analyte. The aim is to select a relatively small number of database samples with similar spectra and similar analyte values to the unknown. These similar samples are then used to make a bias adjustment to the prediction for the unknown using the global calibration. The approach is relatively simple, but relies on having a reasonably good global calibration as a starting point.

40.3 CARNAC-D

The original version of CARNAC aroused very little interest and eventually the programs became obsolete. In 2000 a relatively new method of data compression, wavelet transformation (WT) [14], was receiving interest from the NIR community and LOCAL was receiving favorable attention. This seemed to be an ideal time for a second attempt at introducing the CARNAC idea utilizing WT. The first requirement was to determine if any of the several available forms of WT was superior for compression of NIR spectra. A study reported in 2003 [15] recommended the Daubechies external phase wavelet of order 4 (db4) as a good choice for a wide range of NIR spectra. Apart from the introduction of WT, the other modifications that resulted in CARNAC-D (the “D” is an abbreviation of *deux*; the name of the technique is derived from the village called Carnac in Brittany, France; see Appendix) were the addition of quadratic regression for similarity calculations and modification of the outlier detection algorithms. The other innovation was the use of MATLAB (Version 12) for all programming. A paper describing the new method and giving (surprisingly good) results with the small database used for program development and testing has been published recently [16]. Although the program operates without supervision, it is most easily understood by considering it in two stages. The operations in the first stage are shown in Figure 40.1.

A biscuit dough database from earlier work [10] was used for the development and testing of programs. Spectra from 1100 to 2500 nm in 2 nm steps were recorded on 39 biscuit doughs using a Neotec 6350 Research Composition Analyzer Mk1 (Pacific Scientific, Silver Spring, MD, USA), Figure 40.2b. The doughs were mixed in the laboratory to varying recipes, and the constituent

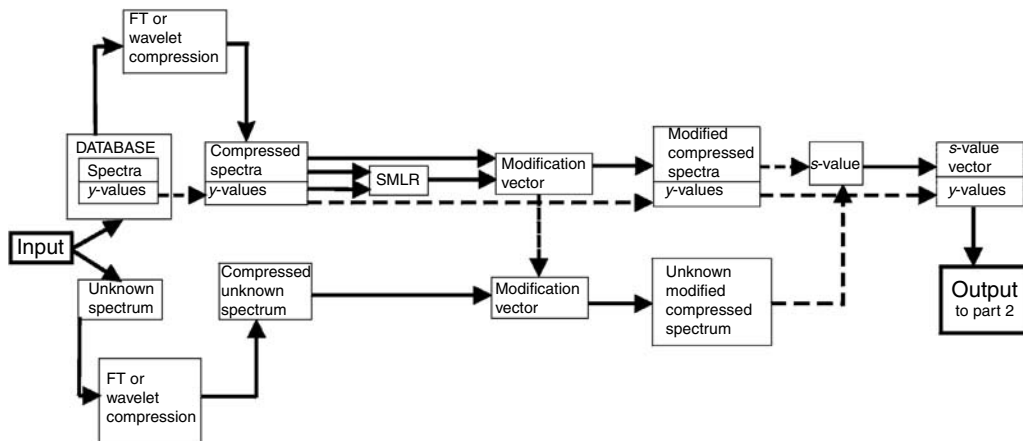


FIGURE 40.1 A flow diagram of CARNAC-D, stage 1. Reprinted with permission of IM Publications from A.M.C. Davies and T. Fearn, *J. Near Infrared Spectrosc.* **14**(6), 403–411 (2006).

predicted is the fat content of the doughs. The spectra were recorded for unrelated work and provided for these studies as ASCII files that were imported into MATLAB (Version 12, The MathWorks, Inc., Natick, MA, USA). The actual type of fat used in the experiment is known in the United Kingdom as “lard”. A spectrum of a sample of lard was measured in 2006 by Dr. Ian Wesley of BRI Australia Ltd on a model 6500 spectrometer (Foss/NIRSystems, Laurel, MD, USA), Figure 40.2a.

40.3.1 COMPRESSION METHODS

Data compression has always been an essential part of CARNAC. It provides two benefits: fewer variables mean concentrated information and faster processing. Transformed spectra are difficult to comprehend visually but they can be retransformed to the wavelength domain. Whether we use Fourier or wavelet transforms for data compression, the basic idea is the same. Each spectrum is represented as a weighted sum of mathematical functions, either sine and cosine waves or wavelets. This representation is exact: the original spectrum can be reconstructed from the weights, which are called coefficients in both schemes. If we had to store all the coefficients there would be no data compression, for there are as many coefficients as points in the original spectrum. However, it is usually possible to reconstruct an acceptable approximation to the spectrum using only a fraction, one quarter perhaps, of the coefficients, and the rest need not be stored. The approximate reconstruction is always a little smoother than the original, and the compression will generally have the beneficial side effect of removing noise. The book by Næs, Isaksson, Fearn, and Davies [17] describes and compares both techniques.

40.3.1.1 Fourier Transformation (FT)

The Fourier decomposition of a signal or a spectrum as a sum of sine and cosine waves of different frequencies has been described in detail (Chapter 6). The fast FT is an available function in MATLAB.

40.3.1.2 Wavelet Transformation (WT)

Wavelets are rather more complex than the use of sine and cosine waves in FT but for understanding CARNAC they can be thought of as a rather special sort of FT. The MATLAB Wavelet toolbox (Version 1.2) was used for WT, and the db4 wavelet was used in the analysis of the biscuit doughs.

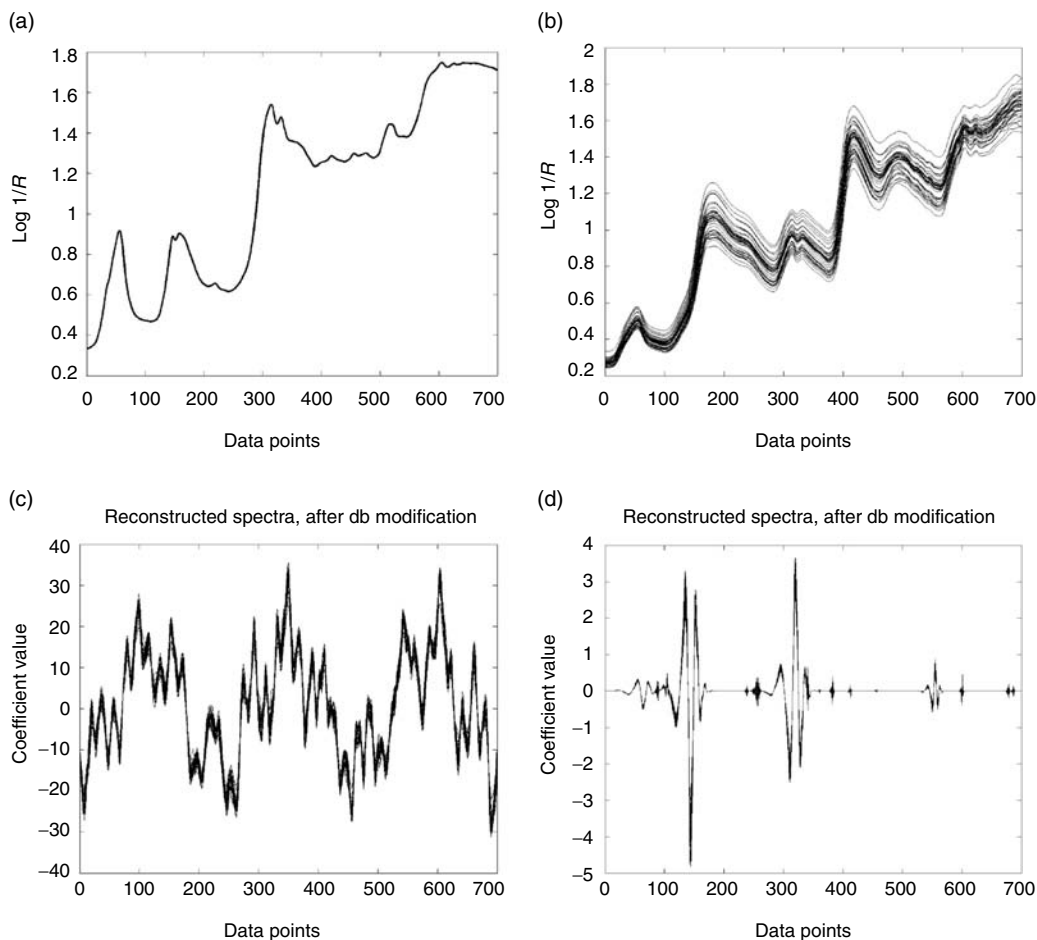


FIGURE 40.2 (a) Spectrum of lard; (b) the original database; (c) the modified versions after retransforming the modified FT compressed; and (d) the modified wavelet compressed versions. Reprinted with permission of IM Publications from A.M.C. Davies and T. Fearn, *J. Near Infrared Spectrosc.* **14**(6), 403–411 (2006).

40.3.2 MODIFICATION

The modification step is an essential part of the technique; it is required to emphasize the signals owing to the analyte of interest. Both FT and WT representations can be retransformed to the wavelength domain. Even compression by a factor of 4 makes little visual difference to the spectra. This is not the case after modification and retransformation. Figure 40.2 shows the effect of one modification.

40.3.2.1 Multiple Regression Method

The main method of modification is by multiplication of the compressed spectrum by a vector of coefficients calculated from a stepwise multiple regression analysis (SMLR) of the values of the analyte of interest in the database on the compressed spectral data. This both selects a subset of the variables and weights them according to their importance for predicting the analyte. The SMLR is run after the compression because FT and WT coefficients are essentially uncorrelated, which makes the unsupervised use of SMLR much less problematic.

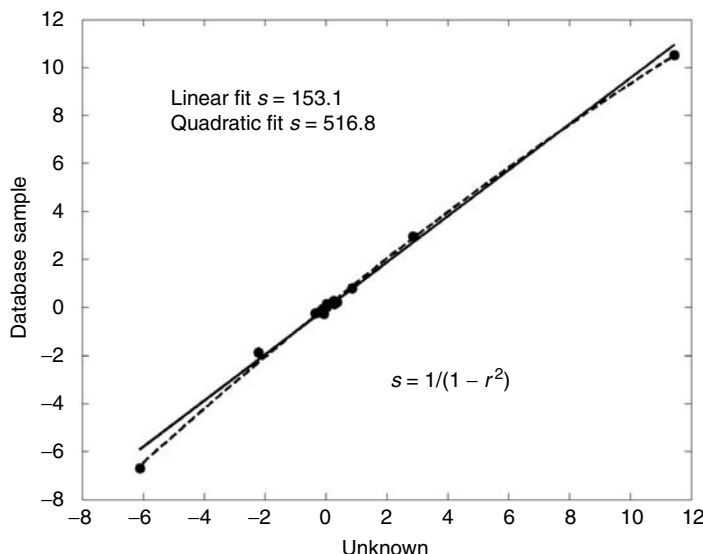


FIGURE 40.3 Correlation plot of the compressed (WT) and modified spectrum of one member of the database against the similarly compressed and modified spectrum of an unknown sample. Circles plotted at the coefficient values for the same data points in each compressed and modified spectrum; solid line, linear fit; dotted line, quadratic fit.

40.3.2.2 Vector of Coefficients Method

As an alternative, the program is able to use an input vector of coefficients that have been derived externally, for example, from the spectrum of the pure analyte. In this case the multiplication is performed before compression.

40.3.3 SIMILARITY

The comparison between an unknown sample and each member of the database is assessed by calculation of a “similarity index” (s) defined by

$$s = 1/(1 - r^2) \quad (40.1)$$

where r is the (simple or multiple) correlation coefficient obtained by fitting either a straight line or a quadratic curve to the relationship between a sample from the compressed and modified database and the unknown sample (after the same compression and modification), Figure 40.3. There are only a few data points in these spectra because of the compression and selection process.

Database samples are selected by their similarity to the unknown. This is done by setting a threshold for s below which samples are ignored, or by selecting the n most similar samples, or by a combination of these strategies.

40.3.3.1 Linear Regression

Linear regression applied to equation (40.1) was the statistic used to assess similarity in the original method [3]

40.3.3.2 Quadratic Regression

The use of quadratic regression to determine r for calculation of s using equation (40.1) was demonstrated to be a useful method for comparing similar samples of differing particle size by Coene,

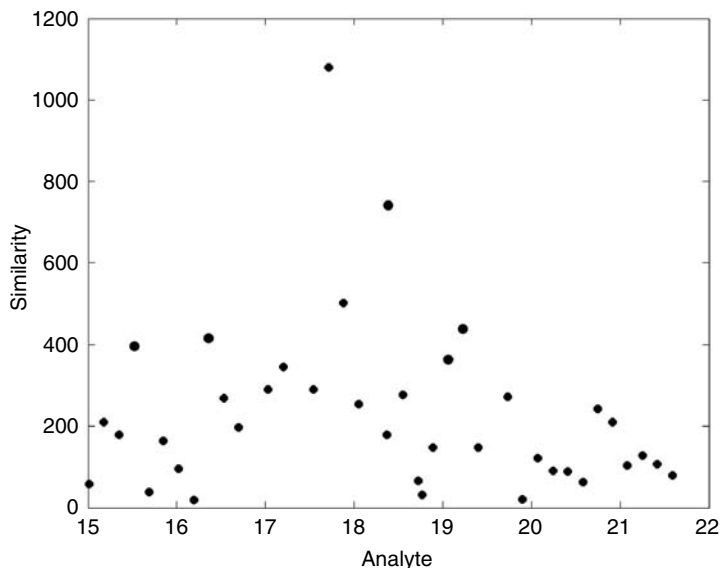


FIGURE 40.4 Plot of similarity against analyte value when testing an unknown sample against each member of a database.

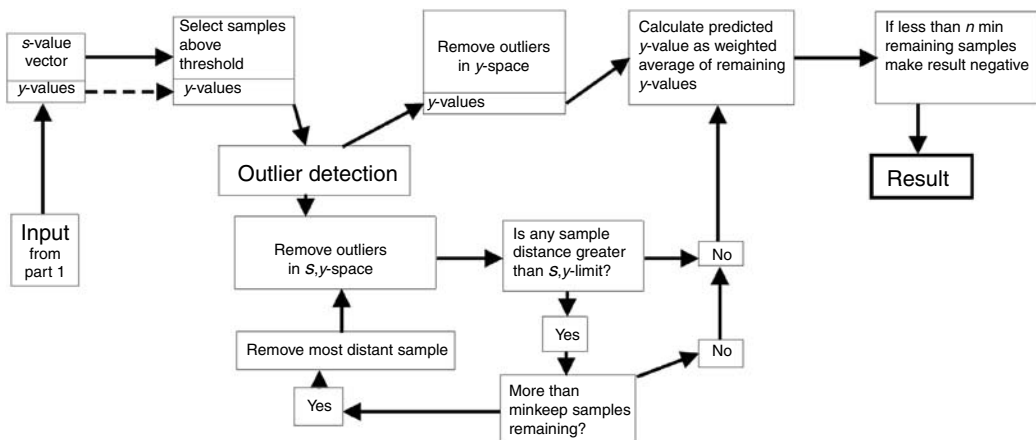


FIGURE 40.5 Flow diagram of CARNAC-D, stage 2. Reprinted with permission of IM Publications from A.M.C. Davies and T. Fearn, *J. Near Infrared Spectrosc.* **14**(6), 403–411 (2006).

Grinter, and Davies [18]. It almost always results in higher values for similarity of very similar samples when compared to linear regression, and this is a useful property for CARNAC.

The output of stage 1 of the program is a plot of s against analyte value, y ; shown in Figure 40.4. The main reason for using s rather than r^2 is that s stretches out the vertical scale in Figure 40.3 in a way that allows one to see the differences between r^2 values of, for example, 0.991 and 0.994, differences that are important in practice. Most of the subsequent calculations, for example, the calculation of the weighted average prediction and the detection of outliers, both described below, use $\log(s)$ rather than s . This measure still emphasizes small differences at the top end of the r^2 scale, but not as much as does s , and works better in the calculations though not in the pictures. The data shown in Figure 40.4 are then automatically processed by the CARNAC-D program according to stage 2 of the flow diagram shown in Figure 40.5, which is optimized by setting various parameters.

40.3.4 OUTLIER TESTS

It may happen that some sample or samples are found that are apparently similar to the unknown sample but are separated from the main group of similar samples. In order to remove these samples, the selected set of samples is tested for the presence of outliers. One of the two methods described below is selected for this function.

40.3.4.1 Quartile Method

This uses only the y -values of the selected samples. The median, quartiles, and interquartile range of these values are calculated, and samples with y -values more than delta times the interquartile range beyond the quartiles are rejected. This is a fairly standard statistical approach to the identification of outliers, with $\delta = 1.5$ being the usual default.

40.3.4.2 Most Distant Samples Method

This method combines information from y and s in assessing distance. The current weighted mean of the y -values for the selected samples is determined using $\tilde{s} = \log(s)$ as weights. The y -values are centered on this prediction and scaled by dividing by a scaling factor f that effectively determines the relative weights given to y and s . The result is a centered and scaled y -value

$$y_c = (y - \hat{y})/f$$

for each of the selected samples, where

$$\hat{y} = \frac{\sum \tilde{s}y}{\sum \tilde{s}}$$

the sums being over the selected samples. Now we “center” the $\tilde{s} = \log(s)$ values by subtracting them from the maximum \tilde{s} value found. This gives, for each selected sample, a value

$$\tilde{s}_c = \tilde{s}_{\max} - \tilde{s}$$

which is zero for the sample with the highest similarity and positive for the others. These two components are combined into a single distance measure d using

$$d = \sqrt{(\hat{y}_c^2 + \tilde{s}_c^2)}$$

If the standard deviation of these distances is greater than a given limit then the most distant sample is removed. This is iterated until the limit is not exceeded or the number of remaining samples is reduced to a prespecified minimum. An example of the available graphic display is shown in Figure 40.6.

40.3.5 DETERMINATION OF WEIGHTED MEAN

The result for a single unknown sample is calculated as a weighted average of the y -values for the remaining selected samples, $\log(s)$ being used as the weighting factors. If the number of selected samples is below the limit for the minimum number of selected samples, then the result is made negative to show that it should be treated as tentative. If no sample is selected, then the result is given as -1.0000 .

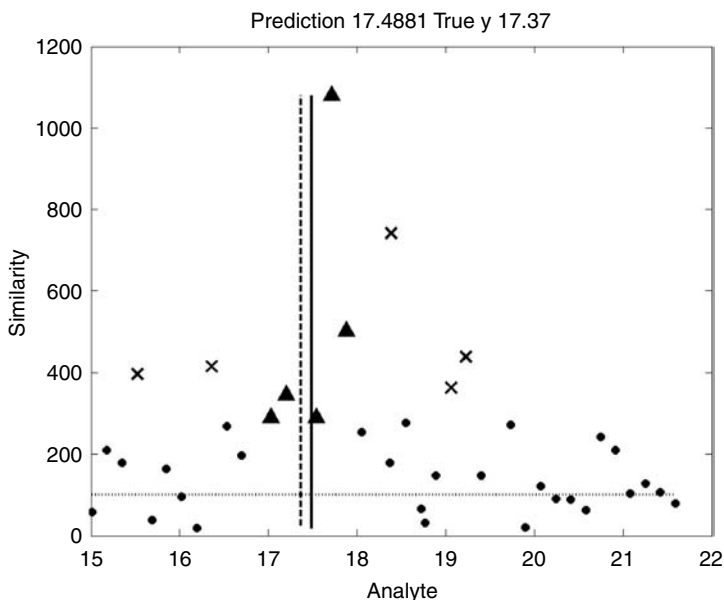


FIGURE 40.6 Automatic processing of the data shown in Figure 40.4 (the plot of similarity for a single sample against the database analyte values).

40.3.6 OPTIMIZATION

The method has currently 22 parameters, about 8 of which need to be optimized for each database and analyte. This is done by trial and error using a test set or cross-validation. With the current software, this involves quite a bit of work. A recent paper [19] describes experiences with three large databases and suggests a strategy for the optimization. Once the optimization has been done, the program should be used for prediction without any alteration of critical parameters.

40.3.7 REPRESENTATIVE RESULTS

During the initialization of the method the program can be set to compute answers for a test set or carry out a cross-validation on the database. The results are used to compute an RMSEP and can also be displayed as a scatter plot.

40.3.7.1 Results from a Small Database

Results from the biscuit dough database obtained by cross-validation are shown in Figure 40.7, by using FT-SMLR processing; Figure 40.8, by using WT-SMLR processing; and Figure 40.9 by using the lard spectrum for vector multiplication followed by WT processing.

40.3.7.2 Results from a Large Database

Dr Gary Ritchie provided sets of spectra of pharmaceutical tablets for the “Software Shootout” at IDRC-2002 and then put all the files into the public domain. This is one of the large databases recently investigated [19]. Figure 40.10 is a plot showing the results of comparing one of the test set samples against all members of the database. The most similar samples form a narrow cluster that was very close to the laboratory value of that sample. Figure 40.11 is the result of testing data from one of the instruments using a database of 460 samples and a test set of 154 samples.

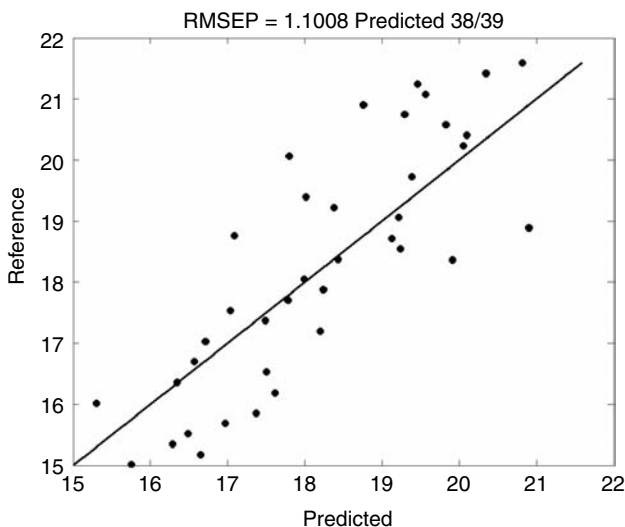


FIGURE 40.7 Results from the biscuit dough database obtained by cross-validation; using FT-SMLR processing.

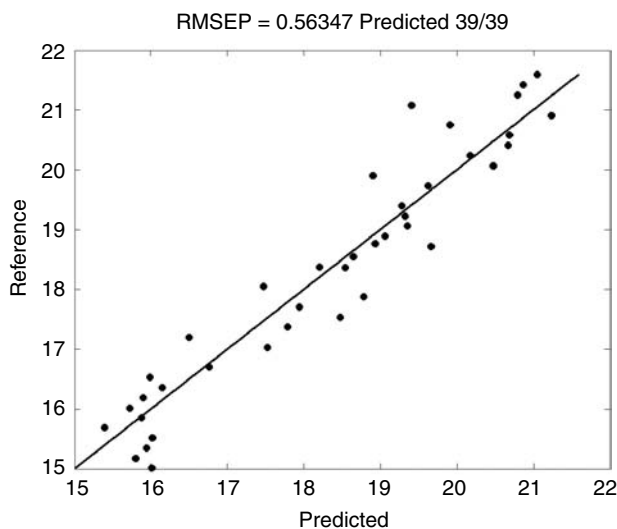


FIGURE 40.8 Results from the biscuit dough database obtained by cross-validation using WT-SMLR processing.

40.3.8 FUTURE PROSPECTS FOR CARNAC-D

We are quite pleased with the results obtained so far. In 2006 we collaborated with Professor Garrido-Vara in a trial of several methods on her very large database of feed samples for IDRC-2006. This study highlighted some problems associated with using CARNAC on this database, in that it contains a large percentage of samples with zero amounts of any particular analyte and these tended to cause difficulties. We are planning some changes to try to improve the performance with this type of data and these may prove useful with more evenly distributed databases. We have had offers of several large databases from various sources for additional testing so we hope that by the end of 2007

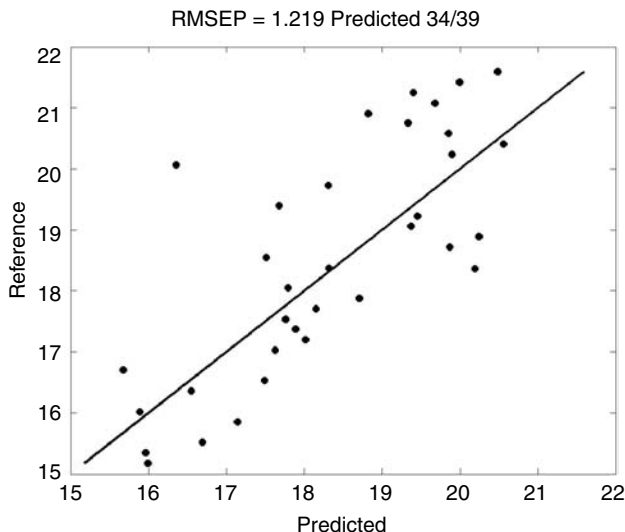


FIGURE 40.9 Results from the biscuit dough database obtained by cross-validation using the lard spectrum for vector multiplication-WT processing.

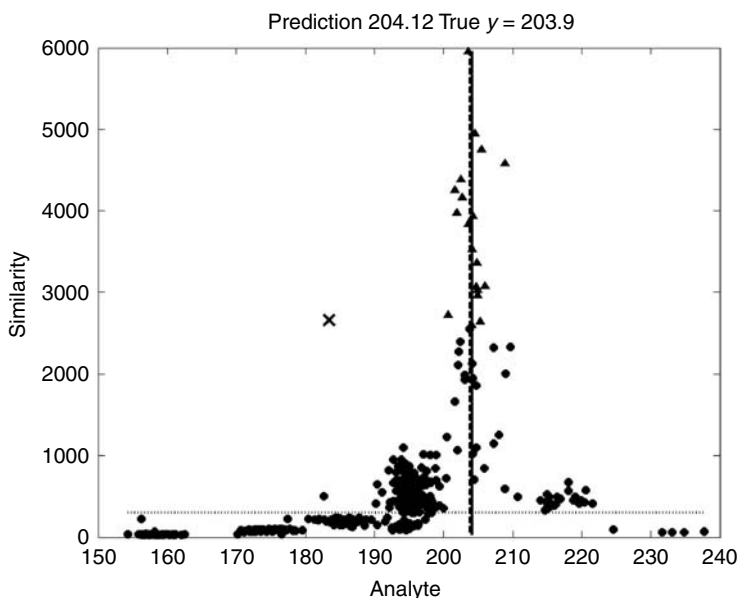


FIGURE 40.10 Typical result of comparing an unknown sample with 460 members of a database. The horizontal dotted line, at 100 is the set threshold value. The black triangles indicate selected samples; crosses indicate samples that were selected but were then rejected by the outlier rejection algorithm. The solid vertical line indicates the predicted value for the test sample while the dashed vertical line indicates the actual database value for the test sample. One outlier was automatically rejected.

CARNAC-D will have been tested with about ten databases. If these are considered sufficiently successful we would hope that it will be made commercially available. In the meantime, the present program is not considered suitable for dissemination but at a later date we might consider limited distribution in a “P code” form of MATLAB.

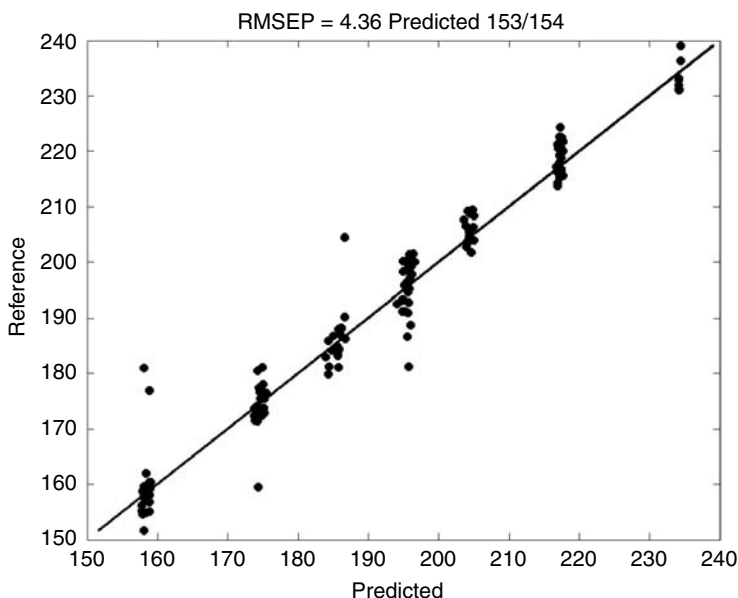


FIGURE 40.11 Result of testing data from one of the instruments using a database of 460 samples and a test set of 154 samples.

APPENDIX: CARNAC AND CARNAC

By Tony Davies

40.A.1 THE LONG HISTORY OF CARNAC, BRITTANY, FRANCE

Officially, the acronym CARNAC stands for Comparison Analysis using Restructured Near-infrared And Constituent data but it really comes from the village in Brittany, France, which is in an area of Neolithic monuments thought to be a gigantic lunar observatory. There are three components in the moon's orbit of the earth and it has been demonstrated that the builders had an understanding of this complexity 3000 years before the "discoveries" of Tycho Braché. The most striking feature, called "alignments," is at Carnac; these are a series of large stones set in 12 rows over a distance of 5 km, Figure 40A.1. It is believed that these were used for extrapolating observations of the moon between periods when observations were prevented by unfavorable weather. I visited Carnac on holiday in 1972 shortly after reading an account of these ideas [21] and was very impressed by the determination and skill of my Celtic ancestors!

40.A.2 A SHORT HISTORY OF CARNAC

In 1983 I was trying to make NIR calibrations on a set of observations that was proving difficult. If I plotted one of the analytes against another it demonstrated many local clusters and I realized that it would take a very long time to obtain samples that would fill the gaps between these clusters and provide reasonable data for an NIR calibration.

In April of 1983, Ian Murray organized an NIR meeting in Aberdeen (which is an 8 h journey by train from Norwich). On the way home I was thinking about this problem, when, for some reason, the plot reminded me of the stones at Carnac (each cluster being a stone). I decided that if I could assign an unknown sample to the correct stone (or cluster) then I would be able to assign analytical values of that cluster to the sample. The extension of this idea was the notion that Neolithic man



FIGURE 40.A.1 The “alignments” at Carnac, France.

may have also used the alignments to improve the precision of their observations to fractions of the distance between stones and thus the idea became one of looking for a set of known samples that were similar to the unknown sample.

The original idea was flawed, because spectra would always be dominated by the major ingredient (or absorber), but it was saved by the concept of modifying the spectra for each analyte. At the time I was collaborating with Fred McClure on applications of FT to NIR spectroscopy so it was natural for me to assume that we would use FT data compression to speed the computations. Working with Fred had also necessitated him to share his CSAS software [22] with me, and so the first version of CARNAC was written as modifications of some of the FORTRAN programs in the CSAS environment. Fred also had some quite large (for 1986) sets of samples. I chose to work first with coffee because one of the analytes, caffeine, was a minor constituent. The first successful run of CARNAC was in Fred’s laboratory at NCSU on July 4, 1986 so we had some extra celebration later in the day! After that I spoke about the idea at IDRC-3 showing results for caffeine in coffee and alkaloids in tobacco. Back in Norwich, we worked with a few additional small databases and I presented a poster at the FT Conference in Vienna in September 1986, which resulted in reference 3.

That was almost the end of CARNAC! Very few people expressed any interest. Although Tom Fearn did include it in his description of “Other calibration techniques” in Chapter 7 of the second edition of *Practical NIR Spectroscopy* [23]. New management at the Food Research Institute closed down my NIR group and made me redundant, and for 19 months I worked on instrument design (the Oxford QN1000) until becoming an independent consultant with even less time for long-term research.

At the IDRC in 1996 there was a lecture about wavelets. I thought it was important but I could not understand it. Later, I talked to the lecturer and he tried to explain it but I could not get it. He said he would send me some papers, which he did, but I did not understand them. About this time, Fred told me that John Shenk had developed a system that sounded a bit like CARNAC and told me that I ought to have another go at developing it. I discussed the idea with Tom Fearn and we agreed that we would do it — sometime! In 1998, Tom Fearn helped me to put on an introductory course at IDRC and he organized an Advanced Chemometrics course. I talked about FT data compression and he described the wavelet compression of the same spectrum. Within 10 min I had understood wavelets!

We agreed that adding wavelets to CARNAC might be interesting. It took until early in 2002 to do the work. The development of CARNAC-D was accelerated by a visit to Australia where I talked about it and we had a very useful discussion with Bob Andersson, Brian Osborne, and Ian Wesley

when we visited Brian's lab in Sydney. By August 2002 we had the working program for CARNAC-D, which I was able to demonstrate at IDRC and we also held an afternoon discussion session on local methods with John Shenk and others. I presented it as an invited speaker at the ICNIRS-2003 conference in Cordoba but it took until the Autumn of 2006 to actually get the paper [16] submitted to JNIRS! Part of this delay was owing to accepting a challenge from Ana Garrido-Varo to take part in a trial of several methods on her very large database of feed samples for IDRC-2006.

It has been slow progress for CARNAC but in the meantime scanning spectrometers, computers, and databases have caught up with the idea. Twenty-four years is a rather extended development time but on the timeline of Carnac it hardly constitutes a scratch on a stone. We are hoping for faster progress in the future!

REFERENCES

1. S.A. Jensen and H. Martens, in *Food Research and Data Analysis*, Ed. by H. Martens and H. Russwurm, Jr., Applied Science, London, p. 253 (1983).
2. R.S. Andersson, B.G. Osborne, and I.J. Wesley, The application of localisation to near infrared calibration and prediction through partial least squares regression. *J. Near Infrared Spectrosc.* **11**, 39–48 (2003).
3. A.M.C. Davies, H.V. Britcher, J.G. Franklin, S.M. Ring, and W.F. McClure, The application of Fourier-transformed NIR spectra to quantitative analysis by comparison of similarity indices (CARNAC). *Mikrochim. Acta (Wien)* **1**, 61 (1988).
4. T. Næs, T. Isaksson, and B. Kowalski, Locally weighted regression and scatter correction for near-infrared reflectance data. *Anal. Chem.* **62**, 664–673 (1990).
5. T. Næs and T. Isaksson, Locally weighted regression in diffuse near infrared transmittance spectroscopy. *Appl. Spectrosc.* **46**, 34–43 (1992).
6. Z. Wang, T. Isaksson, and B.R. Kowalski, New approach for distance measurement in locally weighted regression. *Anal. Chem.* **66**, 249–260 (1994).
7. G. Sinnaeve, P. Dardenne, and R. Agneessens, Global or local? A choice for NIR calibrations in analyses of forage quality. *J. Near Infrared Spectrosc.* **2**, 163 (1994).
8. J.S. Shenk, M.O. Westerhaus, and P. Berzaghi, Investigation of a LOCAL calibration procedure for near infrared instruments. *J. Near Infrared Spectrosc.* **5**, 223–232 (1997).
9. P. Berzaghi, J.S. Shenk, and M.O. Westerhaus, LOCAL prediction with near-infrared multi-product databases. *J. Near Infrared Spectrosc.* **8**, 1–9 (2000).
10. F.E. Barton, II, J.S. Shenk, M.O. Westerhaus, and D.B. Funk, The development of near infrared wheat quality models by locally weighted regressions. *J. Near Infrared Spectrosc.* **8**, 201–208 (2000).
11. P. Dardenne, G. Sinnaeve, and V. Baeten, Multivariate calibration and chemometrics for near infrared spectroscopy: which method? *J. Near Infrared Spectrosc.* **8**(4), 229–237 (2000).
12. R.G. Damberg, D. Cozzolino, W.U. Cynkar, L. Janik, and M. Gishen, The determination of red grape quality parameters using the LOCAL algorithm. *J. Near Infrared Spectrosc.* **14**(2), 71–79 (2006).
13. T. Fearn and A.M.C. Davies, Locally-biased regression. *J. Near Infrared Spectrosc.* **11**, 467 (2003).
14. I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA (1992).
15. T. Fearn and A.M.C. Davies, A comparison of Fourier and wavelet transforms in the processing of near infrared spectroscopic data: Part 1. Data compression. *J. Near Infrared Spectrosc.* **11**, 3 (2003).
16. A.M.C. Davies and T. Fearn, Quantitative analysis via near infrared databases: Comparison analysis using restructured near infrared and constituent data-deux (CARNAC-D). *J. Near Infrared Spectrosc.* **14**(6), 403–411 (2006).
17. T. Næs, T. Isaksson, T. Fearn, and T. Davies, *A User-Friendly Guide to Multivariate Calibration and Classification*, NIR Publications, Chichester, p. 71 (2002).
18. M.P.D. Coene, R. Grinter, and A.M.C. Davies, The use of quadratic regression in qualitative near infrared and visible spectroscopic analysis. *J. Near Infrared Spectrosc.* **4**(1), 153–161 (1996).
19. A.M.C. Davies, T. Fearn, C. Paul, B.G. Osborne, and I.J. Wesley, Quantitative analysis via near infrared databases: application of comparison analysis using restructured near infrared and constituent data-deux (CARNAC-D) to three databases. Submitted to *J. Near Infrared Spectrosc.* (2007).

20. B.G. Osborne, T. Fearn, A.R. Miller, and S. Douglas, Application of near infrared reflectance spectroscopy to the compositional analysis of biscuits and biscuit dough. *J. Sci. Food Agric.* **35**, 99 (1984).
21. S. Mitton, *New Scientist* **13**(April), 60 (1972).
22. A. Hamid and W.F. McClure, Software for an on-line computerized spectrometer. NCARS Bulletin No. 252, North Carolina State University, Raleigh, NC (1978).
23. B.G. Osborne, T.Fearn, and P.H. Hindle, in *Practical NIR Spectroscopy: With Applications in Food and Beverage Analysis*, Longman Scientific & Technical, Harlow, Essex, p. 134 (1993).

Index

A

- AAA (abdominal aortic aneurysm), 657–658, 665, 669, 670
- Absorption techniques, 23
- absorbance to concentration, relating, 126–128
 - of MIR and NIR, 9–13
 - anharmonic oscillator, 11–12
 - harmonic oscillator, 9–11
 - specific characteristics of, 11
- Acetylsalicylic acid, NIR spectra, 8
- Additive or nutrient premixes, NIR analysis of, 405–406
- Adulteration, 247–248, 252, 254, 597, 619, 627, 628, 772
- Agricultural products, NIR application to, 347–386
- alcoholic hydroxyl group
 - vibrations and absorptions of, 350
 - basic characterizing wavelengths, 356–357
 - calibrating NIR instrumentation, 369–374
 - compiling a sample/spectral library, 369–370
 - global calibrations, 371
 - sample selection, 371
 - local calibration, 372–373
 - mathematical transformations of spectral data, 373–374
 - NIRS algorithms, 373
 - quantification, 373
 - repeatability file, 370–371
 - samples with spectral features, 372–373
 - selecting calibration equations, 373–374
- calibration transfer, 375–380
- basic problem defined, 375
 - factors affecting, 376
 - calibration, 376
 - electronics, 376
 - optics, 376
 - sample, 376
 - methods of transfer, 375–377
- calibration transfer goal, 378–379
- chemical bonds and NIR absorption, 349–351
- derivatized spectra, 352–353
- instrument standardization, new methods, 377–378
- monitoring, 380–382
- action, 382
 - control limits, 380–381
 - recalibration, 382
 - standard error of a difference, 380
- 1968 to present, 348–349
- particle size effects in forage materials, 360
- reference methods for, 360–369
- routine agricultural application, 374–375
- specific functional groups bands in, 353–355
- theoretical aspects, 349–360
- water and hydrogen bonding, 355–360
- Agro-forestry products, NIR in verifying label information in, 387–398
- animal feed industry, issues in, 390–392
 - labeling of Spanish food products, 387–398
 - traceability, labeling, and certification of agro-forestry systems, 392–395
- AI (artificial intelligence) analysis, using FT, 113–116
- instrument anomalies in real time, testing for, 115–116
 - instrument noise in real time, checking, 115
 - spectral searching, matching, and component identification, 113–115
- Albumin, 675
- Alcohols, 545–551
- alcoholic beverages, NIR analysis of, 458–459
 - beers, 458
 - wines and distilled liquors, 458–459
 - alkoxy alcohols, 548–550
 - analyzing methods, 551
 - esterified alcohols, 548
 - ethoxylates, 548
 - fatty alcohols, 545–547
 - polyhydric alcohols, 547–548
 - sulfonated alcohols, 550
 - synthetic alcohols, 547
- Alkyds, 551–552
- Allylisopropylacetureide, 586
- Amino acids, 362, 363, 622, 623
- N*-aminoethylpiperazine (AEP), 556
- Amiodarone, 598
- Ammonium lauryl sulfate, 550
- Amphetamine, 635, 636, 644
- Analytical chemistry, NIR in, 232
- Aneurysm, 657, 658, 663, 664
- Angioscopy, 664
- Anharmoncities, 15, 16, 441, 543
- calculation, 12
 - oscillator, 11–12
- Animal feed industry, labeling issues in, 390–392
- Anisotropic scatter effect, 37–41
- Anomalous dispersion, 42
- ANOVA (analysis of variance), 163, 177, 379
- Aorta spectrum, 662–663
- AOTF (acousto optical tunable filter), 226, 526, 534, 596, 707, 743, 745
- principle of operation, 744
- AOTS (acousto-optical tunable spectrometers), 534, 547, 557, 745
- API (active pharmaceutical ingredient), 591, 592, 595, 599, 603, 748
- Archimedian screw, 747
- Aromatic polyesters, 552–553
- Ash, NIR analysis of, 425–426

Association theory vs. macromolecular theory, 530
 ASTM practice for multivariate regression, 148–149
 Atherosclerosis, 657, 663, 664, 665, 670
 At-line/on-line analyses, 512–519
 of carpet yarn heatset temperature, 513–515
 at-line/'pseudo' on-line measurement
 of tire and carpet yarn moisture and finish-on-fiber,
 516–519
 on-line measurement of PVA size on warp yarns,
 515–516
 ATR (attenuated total reflectance), 16, 523, 553
 ATRP (atom transfer radical polymerization), 541
 Autoclave process, 501
 Autoscaling, 208
 Auxiliary statistics from regression, 159–161
 F for regression, 160–161
 multiple correlation coefficient, 160
 standard error of estimate, 160
 Student's *t* for the coefficients, 161

B

Back-scattered light, 24
 Backward and forward scatter, 24
 BAN (barium activity number), 490–491, 493
 Basic principles, of NIR, 7–19
 and Raman and MIR, comparison
 basic instrumentation, 16–18
 current monochromator/detection principles, 17
 process-control aspects, 17–18
 qualitative and quantitative aspects, 15–16
 of vibrational spectroscopy, 9–18
 Beamsplitter, 81, 83
 BEAST, *see* Quantile BEAST algorithm
 Beer's law, 126–128, 137, 152–153
 Beer–Lambert law, 42
 inverse Beer's law, 153
 Beers, 458
 Benford's equations, 45
 Beverages, NIR analysis of, 457–463
 alcoholic beverages, 458–459
 formula products, 460–462
 liquid cell design, 457–458
 nonalcoholic beverages, 459–460
 sample preparation, 457
 Bilinear modeling, 193–195
 Biomedical applications, 647–653
 of near-infrared spectroscopy, 647–653
 blood glucose, 647–648
 blood oxygen, 648–650
 tissue, 650–653
 bio chemistry, 652
 cancer, 653
 chemistry, 652
 Biscuits, NIR analysis of, 409
 Bisphenyldiglycidyl ether, 556
 Bitumen analysis, 525
 Blood glucose determination, 647–648, 676–697
 wavelength interval selection, 676–681
 changeable size moving window partial least
 squares, 678–679
 moving window partial least-squares regression,
 677–678

 searching combination moving window partial
 least squares, 679–686
 noninvasive blood glucose assay, application to,
 686–689
 Blood oxygen, 648–650
 Bouguer–Lambert law, 42, 48
 Bowling lane condition, 569–578
 lane nomenclature and surfaces, 570
 methods, 573–576
 instrumentation, 573–574
 materials, 573
 measurements, 574–576
 calibration, 574–576
 D mapping, 576
 near-infrared reflectance spectroscopy, using 569–578
 NIR-measured lane, 578
 NIR spectra, 576–577
 oil patterns, 570–572
 op-trac, 572
 results, 576–578
 entire lane, mapping, 577–578
 the “tape reader,” 572
 Boxcar smoothing, 104
 Bread, NIR analysis of, 408–409
 Breakfast cereals, NIR analysis of, 410
 Bubble-free liquid cell, 733
 Butter, NIR analysis of, 428–429
 Bypass filters, 733–734

C

Caffeine, 793
 clusters, 620
 Calibration basics, NIR, 123–150
 absorbance to concentration, relating, 126–128
 calibration error
 sources and recommended action, 131–132
 calibration process, 132–140
 calibration set, selection, 132–134
 mathematical model
 curve fitting technique, 138
 developing, 135–137
 difference math technique, 138
 difference ratios technique, 138
 factor analysis (FA)
 FT-NIR technique, 138
 independent variable terms, 136
 log R technique, 138
 MLR technique, 138
 PCA technique, 138
 PLS technique, 138
 standard methods of validating, 137
 technique, 138
 validating, 137–139
 optical data, collecting, 134–135
 routine analysis and modeling, 139–140
 standard concentrations, determination, 134
 statistical tests, 139
 transfer of calibrations, 140
 calibration times reduction, using Fourier transforms,
 112
 laboratory error, 130
 major error sources in NIRS, 128–130

- meaning of outliers, 130
- NIR sensitivity and detection limits, 126
- packing error, 130
- population error, 130
- statistical terms, 140–141
 - mean square for regression (MS_{regr}), 141
 - mean square for residuals (MS_{res}), 141
 - sum of squares for residuals (SS_{res}), 141
 - sum of squares regression (SS_{regr}), 140
- test statistics, 141–149
 - ASTM practice for multivariate regression, 148–149
 - bias-corrected standard error, 145–146
 - coefficient of multiple determination, 142–143
 - F-test statistics for regression, 141–142
 - offset sensitivity, 146
 - partial F or t^2 test for a regression coefficient, 144
 - principal components regression and partial least squares testing, 147–148
 - random variation sensitivity, 146
 - SDD, 146
 - SEE, 144–145
 - SEL for wet chemical methods, 147
 - standard error of cross-validation (SECV), 145
 - standard error of prediction, 145
 - Student's t value, 143
 - validation and monitoring chart, 125
 - and validation of process sensors, 245–263
- Cancer, 653
- Carbohydrates, 62, 364, 459, 487, 623
- Carboxylic acid anhydride, 558
- Cardiovascular disease, 657–670
 - near-infrared spectrometry in, 657–670
 - plaque complication, 664
 - vulnerable conditions, 664
- CARNAC-D, 783–790
 - compression methods, 784
 - flow diagram, 784
 - Fourier transformation, 784
 - modification, 785–786
 - coefficients method, vector of, 786
 - multiple regression method, 785–786
 - optimization, 789
 - outlier tests, 788
 - most distant samples method, 788
 - quartile method, 788
 - representative results, 789
 - similarity, 786–787
 - linear regression, 786
 - quadratic regression, 786–787
 - wavelet transformation, 784–785
 - weighted mean determination, 788
- Carpet fiber identification for recycling, 508–511
- Casein and caseinates, NIR analysis of, 427–428
- Caverta, 641
- Cellulose, 363, 486, 491, 494, 530
 - chemical reference methods for, 364
 - NIR spectra, 8, 402
- Cellulose ester, 544
- Cereal products, NIR analysis of, 399–413
 - additive or nutrient premixes, 405–406
 - packaging materials, 406
 - cereal foods, 407–411
 - biscuits, 409
 - bread, 408–409
 - breakfast cereals, 410
 - extrusion cooking, 410–411
 - flour confectionery, 409–410
 - pasta, 410
 - dough, 406–407
 - flour, 399–400
 - moisture, 401–405
 - ash, 402
 - baking quality of flour, 404–405
 - color, 402–403
 - particle size, 401–402
 - starch damage and water absorption, 403–404
 - wheat gluten, 405
 - protein, 400–401
- cGMP (current good manufacturing practices), 247–248, 254, 255, 258, 581
- Cheese, NIR analysis of, 429–434
 - effect on precision, 430
 - major constituents, 431
 - minor constituents, 431–432
 - ripening time determination, 432–434
 - sample preparation, 430–431
- Chemical bonds and NIR absorption, 349–351
- Chemical reference methods, for agricultural products, 362–367
 - carbohydrates, 364
 - cellulose, 364
 - fiber, 363–365
 - in vitro and animal response, 366–367
 - minerals, 365–366
 - moisture, 362
 - noncellulose polysaccharides, 364
 - protein, 362–363
- Chemometrics, 126, 207, 297, 632–633, 673–697
 - interference signals removal, 692–697
 - OSC vs. ROSC-pretreated spectra, 694
 - pretreated spectra, 693–694
 - spectral regions selection, 694–697
 - region orthogonal signal correction, 689–692
 - orthogonal signal correction, 690–691
- Chloroform, 745
 - NIR spectrum of, 745
- Chondroitin sulfate, 627
- CLS (classical least squares) models
 - advantages and disadvantages, 214
 - ILS calibration versus, 214
 - paradigms of, 214
- COA (certificate of analysis), 621
- Collagen, 658–659, 661
 - NIR spectrum of, 659, 663, 667
- Combination bands, 12, 13, 16, 62, 67, 488–489, 532, 536, 541, 543, 555
- Commercial NIR instrumentation, 67–78
 - basic instrument configurations, 71; *see also* Filters, in NIR instrumentation
 - common commercial UV-Vis-NIR, NIR, and MID-IR instrumentation, 75–77
 - dispersive (grating-type) scanning NIR instruments, 73–74
 - distinguishing characteristics of, 69
 - instrument specifications for comparison, 70
 - instrumentation design, 68–75
 - single monochromator scanning instrument, 74

Composition estimation, using Fourier transforms, 109–111

Computer storage requirements reduction, using Fourier transforms, 111–112

Computing derivatives via Fourier space, 106–109
by convolution, 106
corrects for particle size anomaly, 109
via the FD, 106
Fourier derivatives, 108–109
by polynomial regression, 106
Taylor series, 106–107

Conditioning system, 735–736

Confidence interval, 165

Connes advantage, of FT-NIR, 84

Continuous processes, 501, 717

Continuum theories of diffuse reflection, 24, 26–41
diffusion theory, 36–37
discrete ordinate approximation, 34–36
K–M theory, 28–34
Schuster's theory, 27–28
two-flux treatments, 26–27

Convolution smoothing in wavelength space, 100–104

Counterfeit cigars, 775–778
Cuban group vs. non-Cubans, 776–777
experimental, 775–778

Counterfeit currency, detection of, 761–768
background, 761–762
best wavelengths identification, 762–768
challenges, 763–768
experimentation, 762

CPAC (Center for Process Analytical Chemistry), 257

Curve fitting technique, 57, 138, 176, 543

CVF (circular variable wavelength filter), 703–704

D

D mapping, 576

Dahm Equation, 44–45, 57

Dairy products, NIR analysis of, 415–437
butter, 428–429
calibration procedure, 417–418
casein and caseinates, 427–428
cheese, 429–434
milk, 418–420
milk powders, 420–427
NIRS in the dairy industry, 416–417
standard methods, 417

Darling–Dennison resonance, 13

Deconvolution via Fourier space, 104–106

Derivatized spectra, of agricultural products, 352–353

Desketoprofen trometamol, 591

Detectors, in FT-NIR, 82
detection limits and sensitivity, NIR, 126

Detrending, 360

Dicumyl peroxide, 535

Dietary supplement industry, 613–628
adulteration, detection of, 627
chondroitin sulfate, 627
dietary supplement testing, challenges of, 614–616
qualitative and quantitative analysis, NIR for, 613–628
mid-IR vs. NIR, 616–617

NIR calibrations, development of, 617–621

quantitative analysis, 624–627
raw root powder vs. extract, differentiation of, 623–624
typical installation, 621–623

Difference math technique, 138

Diffuse reflection
absorption and scatter by a single particle, 23–26
Mie scattering, 24–26
application of theory to model systems, 48–59
continuum theories of, *see* Continuum theories of diffuse reflection
diffuse vs. directed radiation, 22–23
Lambert cosine law, 22–23
discontinuum theories of, 41–48
reflection measurements for, 59–62

Diffuse transmission, 24

Diglycidyl ether dimethacrylate, 560

Dioctyl phthalate, 536

Dipyrone, 639

Discontinuum theories, of diffuse reflection, 24, 41–48
representative layer theory, 46–48
theory for an assembly of spheres, 42
theory for sheets and an assembly thereof, 42–45

Discrete ordinate approximation, 34–36

Discriminant analysis, 297

χ^2 distribution, 165

DMA (dynamic mechanical analysis), 542

Dough, NIR analysis of, 406–407

DR (digital regression), 225
OR versus, 226–227

DSC (differential scanning calorimetry), 534

DTGS (deuterated triglycine sulfate), 523

Dyed nylon carpet yarns, 505

E

Elastin, 658, 661, 662
NIR spectrum of, 660

Elastography, 664

EPDM (ethylene propylene diene monomer), 536

Epoxy resins, 553–560, 772

ERT (equation of radiative transfer), 26, 59

ESR (electron spin resonance) spectroscopy, 535

ETBE (ethyl *tert*-butyl ether), 522

Ethanol, 190, 522

Ethylene glycol diacrylate, 542

Euclidean distance measures, 308–327

EVOH (ethylene-vinyl alcohol copolymer), 547

Extended spectrum concept, 3

Extinction, 24

Extinction coefficient, 67, 156, 161, 163

Extrusion cooking, NIR analysis of, 410–411

F

Factor analysis (FA) technique, 138

Fat, NIR analysis of, 424

Fermi resonance, 13

Fiber
chemical reference methods for, 363–365
fiber/polymer-type identification, 507–508

Fiber-optic light guides, 741

FID (flame ionization detection), 522

Filters in NIR instrumentation

- interference filter instrument, 72
- Neotec filter-type NIR spectrophotometers, 72
- tilting filter concept, 72–73
- turret-mounted discrete filter instrument, 73
- wedge interference filters, 72

Finish-on-fiber measurements, 499–501

Flour confectionery, NIR analysis of, 399–400, 409–410

Forage spectrum, of agricultural products, 352–354

Formaldehyde, 562–563

Formula products, 460–462

- accuracy and precision, 461–462
- milk-base formula calibration for major constituents, 462

Foss-NIRSystems Model 6500, 469

Fourier transform spectrophotometers in the NIR,
see FT-NIR

Fourier transforms, analysis using, 93–121, 784

- artificial intelligence, 113–116
- Fourier mathematics, 94–95
- interferogram space, 117–119
- periodicity requirement, 95–97
- qualitative spectral analysis, 100–109
- quantitative spectral analysis, 109–113
- square-wave spectrum, 95
- Fourier coefficients of, 96

Fraunhofer diffraction, 25

Fresnel formulas, 129

Front-surface reflection effect, 38

Fruit juices, 459–460

FSD (Fourier self-deconvolution), 101

FT/IR analyses, 479–481; *see also* Lignocellulose study

FT-NIR (Fourier Transform Near-Infrared)

- spectroscopy/spectrophotometers, 79–91, 138, 615
- advantages of, 83–84
 - Connes advantage, 84
 - constant resolution, 84
 - continuous spectra, 84
 - multiplex advantage (Fellgett advantage), 83–84
 - negligible stray light, 84
 - throughput advantage (Jacquinot advantage), 84
- common sampling options, 80
- Fourier transform interferometry, 82–84
- FT-NIR spectrophotometer design, 80–82
 - beam splitter, 81
 - detectors, 82
 - interferometer, 81
 - laser, 81–82
 - optical components, 82
 - source, 81
- optical layout of, 80
- quantitative NIR measurements, requirements, 84–88
- standards, 88–91

Fuels, 521–522

- composition of, 524–525
- oxygenates in, 522–523

FWHH (full width at half-height), 101

G

Gas chromatography, 522

Gases, NIR analysis of, 439–456

- aromatic gases, 442
- chlorinated alkanes, 442
- MIR and NIR spectra, comparison, 440–441
- mix-match search algorithm, 450
- natural gas compositions and BTU contents, 443–448
- NIR gas-phase library and search with mixtures, 448–455
 - search algorithm, 449–450
 - search results, 450–454
 - and solutions spectra, comparison, 441–443
 - spectral library, preparation, 448–449
 - substitution patterns in, 441
- Gauss–Jordan method, 325–326
- Geometric optics, 24
- γ -Globulin
 - NIR spectra, 683
 - SCMWPLS, determination with, 681–686
- Glovebox, 710–715
 - NIR tracings, 712
 - sample holder, 712
- Glucose in water solutions, 460
- Gram–Schmidt orthogonalization, 176
- Granulation, 595
- Greige nylon carpet yarns, 503–505

H

Hank process, *see* Autoclave

Harmonic oscillator, 9–11

HDL (high-density lipoprotein), 658

HDPE (high density polyethylene), 534, 702

Heatset temperature measurements, 501–503

- in textiles, 503–505
 - dyed nylon carpet yarns, 505
 - Greige nylon carpet yarns, 503–505

Hexamethylenetetramine, 561

Hidden mass effect, 36

Honeydew, 486

Honigs reconstruction, 336–337

HPLC (high performance liquid chromatography), 536, 589

Human serum albumin

- SCMWPLS, determination with, 681–686

Hyperspectral imaging, 605

I

ILS (inverse least squares) calibration, 214

In vitro and animal response, chemical reference

- methods for, 366–367

Indicator variables in NIR analysis, 166, 297–306

- data manipulation, 301–302
- interpretation, 299–301
- problem, 298–299
 - solution, 303
- procedure, 299
 - all possible-combinations (APC) search, 299
 - traditional step-up search, 299

InfraAlyzer 400R, 468

Infrared plant analyzer, 718

Instrument recalibration, in multivariate calibration

- models, 241–242

Intact sampling, 532–533
 Interrelation, effect of, 161–162
 Interference filter instrument, 72
 Interferogram space analysis, using Fourier transforms, 117–119
 chemistry estimation from, 117–118
 no-moving-parts FT-NIR, 118–119
 Interferometry, Fourier transform, 82–84
 interferometer, in FT-NIR, 81
 Michelson interferometer, 82–83
 principle, 82
 Inverse Beer's law, 153
 IPN (interpenetrating polymer networks), 560
 Isobestic point, 554
 IVUS (intravascular ultrasonography), 664

J

Jacobi polynomials, 35

K

Karl Fischer method, 134, 361
 KES/WRONZ NiraSpec, 469
 KFR (Karl Fischer Reagent), 498
 Kjeldahl method, 404
 K–M theory, 28–34, 58, 129
 anisotropic scatter effect and, 37–41
 deviations from, 37–41
 variables used in, 29
 Kortüm's notation, 31
 Kubelka–Munk (K–M) theory, *see* K–M theory

L

Laboratory error, 130
 Lactose, NIR analysis of, 424–425
 Lambert cosine law, 22–23, 42, 129
 Laser, in FT-NIR, 81–82
 Latent variables, 179
 Layer representative of particulate solids, model for, 46
 LDL (low-density lipoprotein), 658
 LDPE (low-density polyethylene), 702
 Lignocellulose study, FT/IR vs. NIR, 479–484
 FT/IR analyses, 479–481
 NIR analysis, 481–483
 Linear absorption coefficient, 42
 Linear regression, 786
 β -Lipoprotein, 674
 Liquids, sampling of, 279–280
 Liquids analysis, 722–745
 influencing factors, 725–745
 fiber-optic-based analyzers, 740–745
 companies offering, 742
 filtering, 731–734
 bypass filter, 731
 pathlength variation, 727–729
 process liquid cells, 734–740
 sample transport systems, 725–727
 bypass sampling system, 727
 temperature variations effect, 729–731

 optimal pathlength, 723–724
 pretreatment functions, 729
 LLDPE (linear low-density polyethylene), 535
 Local methods, 780–783
 CARNAC, 782
 LOCAL, 783
 locally biased regression, 783
 locally weighted regression, 782–783
 Local mode concept, 13
 — log R technique, 138
 Lubricant, 499–500, 526, 561
 LWR (locally weighted regression), 218

M

MA (methyl acrylate), 541
 Macromolecular theory vs. association theory, 530
 Mahalanobis distance, 130, 312–314, 323, 617, 714
 advantage, 315, 330
 limitation, 330
 for quantitative analysis, 324
 Mathematical expression of representative layer, 47–48
 Mathematical model development
 in NIR calibration, 135–137; *see also under*
 Calibration basics
 Matrix equation, 312
 Matrix referencing, 52–55, 61–62
 MBP (methyl 2-bromopropionate), 541
 MCFT, 118
 Mean centering, 208–210
 Medullation, 472
 Melamed's theory/model, 24, 42
 Mercerization, 490–494
 Methacrylic polymers, 541–543
 Metric spaces, 318
 Michelson interferometer, 82–83
 Microprocessor, 5
 Microwave drying, 367–368
 Mie theory/scattering, 24–26, 35–36
 Milk, NIR analysis of, 418–420
 milk powders, 420–427
 ash, 425–426
 fat, 424
 general, 420–421
 protein, 421–422
 lactate, 425
 lactose, 424–425
 moisture, 422–423
 repeatability and reproducibility, 426–427
 skim milk powder content, 426
 Milling, 589–590, 593, 595
 Minerals, chemical reference methods for, 365–366
 MIR spectroscopy, absorption techniques of, 9–13
 MLR (multiple linear regression), 138, 215, 297, 500, 523, 534, 590
 MMA (methyl methacrylate), 534, 541
 Moisture register NIR analyzer, 755
 Moisture
 chemical reference methods for, 362
 NIR analysis of, 401–405, 422–423; *see also under*
 Cereal products
 MON (motor octane) method, 523
 Monitoring, NIRS analyses, *see under* Agricultural products

- Morphometry, 659
- Moving window partial least-squares regression, 677–678
- MPA (moving point average) smoothing, 100–101, 103
comparison of derivatives obtained by, 108
effect of, 105
- MRI (magnetic resonance imaging), 664
- MSC (multiplicative scatter correction), 210–211, 522
- MSPC (multivariate statistical process control)
processes, 251
- MTBE (methyl *tert*-butyl ether), 522–523
- Multicollinearity, 161
reduction, using Fourier transforms, 113
- Multilinear regression and principal component analysis, 151–188
auxiliary statistics from regression, 159–161
calibration with error, 156
calibration with error in Y only, 156–159
calibration principles, 152–156
error-free case, 152–156
effect of intercorrelation, 161–162
extraneous variables in a calibration, 166
principal component analysis (PCA), 166–178;
see also PCA
wavelength selection
with error in Y only, 163–165
with multiple wavelengths, 163
with noise in the optical data, 164–165
in the single-wavelength case, 163
- Multiplex advantage (Fellgett advantage), of FT-NIR, 83–84
- Multivariate calibration aspects for NIR, 207–229
data preprocessing, 208–214
instrument standardization, 213–214
mean centering, 208–210
multiplicative scatter correction (MSC), 210–211
orthogonal signal correction (OSC), 211–213
variance scaling, 208–210
model selection and validation, 218–221; *see also* Outliers
model selection, 220–221
- multivariate calibration, 214–218
locally weighted regression (LWR), 218
multiple linear regression, 215
partial least squares regression, 216–218
principal components regression, 215–216
optical regression, 224–228
- Multivariate calibration models transfer
based on near-infrared spectroscopy, 231–243
NIR calibration models, development of, 232
steps in, 232
NIR in analytical chemistry, 232
situations yielding invalid predictions, 233
avoiding such situations, 233–234
instrument recalibration, 234, 241–242
robust calibration models, using, 234, 240–241
standardization methods, 233–240
- cotton fiber maturity, 494–496
degree of mercerization, 490–494
draw blending, 489
intimate blending, 489
polyester/cotton blends, quantitative analysis, 489–490
- Natural gas compositions, NIR analysis of, 443–448
NIR field measurements, 447–448
resolution required for quantitative predictions, 446–447
- Near-infrared (NIR) radiation, *see* NIR
- Neave and Wheeler's Shewhart's charts, 252–253
- Neotec filter-type NIR spectrophotometers, 72
- NIR (near-infrared radiation) spectroscopy, 521–526, 701–706
basic principles of, *see* Basic principles
biomedical applications, *see under* Biomedical applications
calibration basics, *see under* Calibration basics
in cardiovascular disease, 657–670
in dietary supplement industry, *see under* Dietary supplement industry
digital revolution and, 5–6
microprocessor, 5
principal components analysis (PCA), 5
discovery of, 3–4
filter analyzer, 551–552
first infrared spectra, 4
historical development, 3–6
indicator variables in, 297–306
for labeling issues in the animal feed industry, 390–392
multivariate calibration aspects for, 207–229
PAT effect on, 582–584
petrochemical industry, *see under* Petrochemical industry
pharmaceutical applications, *see under* Pharmaceutical applications
test data, predicted value of, 706
training data, predicted value of, 706
- NN (neural networks), 648
- No-moving-parts FT-NIR, 118–119
- Nonalcoholic beverages, NIR analysis of
fruit juices, 459–460
glucose in water solutions, 460
- Noncellulose polysaccharides, chemical reference
methods for, 364
- Noncontact NIR analyzers, 755–757
- Non-Euclidean distance measures, 308–327
- Norethisterone, 588
- Normalization, 168, 171, 317–318
- NPN (nonprotein nitrogen), 421
- Nylon fibers, moisture content in, 498–499
- Nyquist criteria, 720

O

N

- Natural fibers, NIR analysis of, 486–497
amount of reducing sugar from cotton surface, 486–489

- OCT (optical coherence tomography), 664
- Octane number, 523–524
motor octane (MON)
research octane (RON) method, 523
- Oil concentrations, NIR-predicted, 577
- Oil patterns, 570–572

Optical analyzer, 737
 Optical components, in FT-NIR, 82
 Op-Trac, 572
 OR (optical regression), in NIR, 224–228
 application, 226–228
 theory, 225–226
 OSC (orthogonal signal correction), 211–213, 241
 Outliers, 130, 218
 detection, 219–220
 in quantitative analysis, 324–325
 Oven-drying, 367–368
 Overtones, 12, 13, 16, 62, 67

P

Packaging materials, NIR analysis of, 406
 Packing error, 130
 Partial least-squares regression, *see* PLSR
 Particle size effects, 109, 161
 PAsT (process assessment technology), 249
 Pasta, NIR analysis of, 410
 PAT (process analytical technologies), 249–260,
 581–584
 NIR, effect on, 582–584
 PAT guidance, 582
 in pharmaceutical industry, *see under* Pharmaceutical
 Industry
 PBS (phosphate-buffered saline), 658
 PCA (principal component analysis), 138, 166–178,
 370, 524, 551, 635–644, 659
 advantages, 184–185
 interpretation of calibration equations, 185
 lack of requirement for wavelength selection,
 184–185
 robustness, 185
 calibration, 178–182
 computation flow in, 180
 characteristics of, 182
 comparison with Fourier analysis, 168–176
 data compression, 185–186
 definition and discussion, 167–168
 limitations, 186–187
 qualitative analysis (SIMCA), 185
 spectral reconstruction, 182–184
 uniqueness of, 176–177
 PCCPs (process critical control parameters), 249–254
 PCR (principal components regression), 215–216, 526,
 659, 782
 PCT (process control technology), 249
 Period of idling, 8
 Periodicity requirement, in FTIR, 95–97
 Perkins' method, 487
 PET (polyethylene terephthalate), 702
 Petrochemical industry, 521–526
 near-infrared (NIR) spectroscopy, 521–526
 bitumen analysis, 525
 environmental/health analysis, 525–526
 fuels, 521–522
 composition of, 524–525
 oxygenates in, 522–523
 gasoline properties, 525
 PGI (Protected Geographical Indication), 388
 Pharmaceutical applications
 of near-infrared spectroscopy, 585–605
 ablet actives, determination of, 587–588
 hardness, 601–603 near-infrared imaging, 604–605
 intact dosage forms, analysis of, 596–601
 considerations for, 603–604
 solid dosage formulations, 588–594
 manufacturing, analysis during, 594–596
 in tablets and solid pharmaceutical dosage forms,
 586–587
 Phenacetin, 586
 Phenolic resins, 561
 Phthalic anhydride, 551
 Pitts's solution, 38
 Pitts–Giovannelli equation, 38–40
 Planck's constant, 10
 Plastics analysis, 699–715
 plastic and rubber waste, characterization of, 710–715
 instrumentation, 711–713
 results, 713–715
 resin identification, *see under* Resin
 PLS technique, 138
 PLSR (partial least-squares regression), NIR
 instruments calibration by 189–205, 522–523,
 534, 594
 data analysis method, 191–198
 bilinear modeling, 193–195
 calibration and prediction, 195–196
 PLS regression method, 191–198
 PLS1 algorithm, 198
 PLS1 regression, 195
 two-block predictive PLS regression, 197
 validation and assessment, 196–199
 equipment and software requirements, 190–191
 experimental, 190–19
 laboratory procedure, 191
 notation, 190–203
 results and discussion, 198–203
 PLSR model, 199–203
 residual statistics, 198–199
 PNA (polynuclear aromatic), 526
 Polyamides, 543–544
 Polyester fiber, by different producers, 507, 551–553
 alkyds, 551–552
 aromatic polyesters, 552–553
 saturated aliphatic polyesters, 552
 unsaturated polyesters, 552
 Polymers, NIR analysis of, 529–563
 applications, 535–563
 acrylic and methacrylic polymers, 541–543
 amide-based, 542–543
 ester-based, 541–542
 alcohols, 545; *see also under* Alcohols
 cellulose ester, 544
 epoxy resins, 553–560
 formaldehyde, 562–563
 phenolic resins, 561
 polyamides, 543–544
 polyesters, *see under* Polyester fiber, by different
 producers
 polyethylene and polypropylene, 535–536
 polystyrene, 539–541
 polyurethanes, 551
 polyvinyl acetals, 539
 polyvinyl alcohol, 537–539

polyvinylchloride, 536–537
 silicones, 561–562
 substitution on starch, 544–545
 near-infrared vs. mid-infrared, 530–535
 combination bands, 532
 identity and quality testing, 533–534
 intact sampling, 532–533
 overtones, 531–532
 process analysis, 534–535
 Polynomial smoothing, in Fourier transform, 101–102
 comparison of derivatives obtained by, 108
 effect of, 105
 Polyvinyl acetals, 539
 Polyvinyl alcohol, 537–539
 Population error, 130
 “Postage stamp” patterns, 574
 Powder mix line, 537
 PP (polypropylene), 702
 PRESS (predictive residual error sum of squares), 679
 Process analysis, 717–757
 liquids analysis
 noncontact analysis, 750–757
 influencing factors, 750–752
 noncontact nir analyzers, 755–757
 sample speed variation, 752–755
 requirements, 719–722
 analysis rate, 720–721
 analyzer environment, 719–720
 process analyzers, 721–722
 in process control, 722
 solids, contact analysis of, 745–750
 sampling parameters, 745–747
 solid analyzers, 747–750
 Process liquid cell, 738
 Process sensors
 calibration and validation of, 245–263
 calibration
 under the 21st century initiative, 255–257
 DMAIC model, 259
 future considerations, 257–258
 PAT process map, 256
 PAT process validation, 250
 risk assessment, management, and continuous verification, 251–255
 validation
 under the 21st century initiative, 249–251
 Product identification technique, 136
 Protein
 chemical reference methods for, 362–363
 NIR analysis of, 400–401, 421–422
 PS (polystyrene), 539–541, 702
 PVC (polyvinyl chloride), 536–537, 702
 Pylox gloves, 714
 adsorbed moisture, effect of, 714

Q

Quadratic regression, 786–787
 Qualitative discriminant analysis, 307–331
 applications, 318–319
 applications involving a single group, 321–327
 detection of samples, 319–320

 to quantitative analysis, 324
 transfer of calibrations, 321
 warnings of misclassification, 320–321
 based on selected wavelengths, 308–327
 non-Euclidean distance measures, 308–327
 comparison of methods, 330
 direct spectral matching use in, 328–329
 direction cosines use in, 328
 group mean matrix, 311
 group mean, 311
 matrix equation, 312
 methods based on use of entire spectrum, 327–330
 multidimensional distance metrics, comparison, 318
 normalization, 317–318
 outliers detection in, 324–325
 principal components use in, 329–330
 sample preparation methods, evaluating, 327
 sample selection in, 325–327
 statistical theory of, 314
 wavelength selection, 316
 Qualitative spectral analysis, in Fourier transform, 100–109
 boxcar smoothing, 104
 computing derivatives via Fourier space, 106–109
 convolution smoothing in wavelength space, 100–104
 convolution, 100
 moving point average (MPA) smoothing, 100–101, 103
 deconvolution via Fourier space, 104–106
 Fourier smoothing, 102–104
 polynomial smoothing, 101–102
 Quantile-BEAST (Bootstrap Error-Adjusted Single-sample Technique) algorithm, 590–591, 597–603
 Quantitative NIR measurements, requirements, 84–88
 Quantitative spectral analysis, using Fourier transforms, 109–113
 calibration maintenance, encouraging, 112–113
 calibration times, reducing, 112
 composition estimation, 109–111
 computer storage requirements reduction, 111–112
 multicollinearity, reducing, 113
 Quartile Method, 788

R

Radiation transfer equation, 26, 29
 Raman spectroscopy, 522, 535
 fluorescence and scattering efficiency in, 14
 scattering technique of, 13–15
 Range splitting, 133
 Ravensara oil, 627
 Reducing sugars, 486
 Reference methods for NIRS analysis, 360–369
 chemical reference methods, 362–367
 concept definition, 360–362
 sample preparation methods for, 367–369
 drying, 367–368
 microwave drying, 367
 oven-drying, 367
 grinding, 368
 mixing, 368

Reference methods for NIRS analysis (*continued*)
 packing, 368–369
 sampling, 367
 unground samples, 369
 total nitrogen and total protein content, 361
 Reflection, 23–24
 reflection measurements, in diffuse reflection
 experimental considerations for, 59–62
 depth of penetration, 59–60
 relative reflectance and matrix referencing,
 effect of, 61–62
 Relative reflectance, 61–62
 Remission fraction, 46
 Repeatability and reproducibility, NIR analysis of,
 426–427
 Representative layer theory, 42, 46–48
 absorption and remission of, 46–47
 layer representative of particulate solids,
 model for, 46
 mathematical expression of model, 47–48
 Resin, 699–707
 NIRS and neural networks, identification using,
 699–707
 experimental, 701–703
 field instrumentation, 703–704
 results, 704–707
 artificial neural networks, 705–706
 field instrumentation, 706–707
 spectroscopy, 704–705
 theory, 701
 RMSE (root mean squared error), 221
 RMSEC (root mean squared error of calibration),
 221
 RMSECV (root mean squared error of cross validation),
 221
 RMSEP (root mean square error of prediction),
 196–198, 221, 592
 Robust calibration models, 234
 development of, 240–241
 against expected sources of variations, 240
 against simulated perturbations, 240–241
 based on pretreated spectra, 241
 Roller compaction, 595
 RON (research octane) method, 523
 Rotating multifilter analyzer, 753
 Routine analysis and modeling, in NIR calibration,
 139–140
 Rozenberg equation, 40
 RTQC (Real Time Quality Control), 246, 262

S

Salicylic acid, 598
 SAM (spectral searching and matching), 113–115
 Sampling/sample preparation/sample selection,
 267–295
 of agricultural products, 367–369; *see also under*
 Reference methods, for NIRS analysis
 blending, 281–282, 287
 Diverter sampler, 275
 error and, 272–282
 sample source, 272
 sampling itself, 272–281

dry or low moisture crystalline solids, 280
 forages, fibers, and related materials, 277–279
 grains and similar materials, 274–277
 liquids, 279–280
 nonsampling, 280
 sample collection, 274
 sampling location, 272–274
 semisolids, 280
 flour/grain impeller continuous sampler, 277
 grinders and grinding, 283–287
 factors affecting grinder performance, 284–287
 grinders, types, 283–286
 burr mill, 283
 cutting type, 284
 hammer mill, 284
 impeller mills, 284
 Retsch centrifugal mill, 284
 manual sampling, 276
 sample-associated factors in NIR analysis, 269
 sample cells and presentation, 289–292
 cell cleanup, 290
 static electricity, 291
 undetected moisture, 291–292
 sample handling, 281
 packaging, 281
 transportation, 281
 sample identification, 281–282
 sample preparation, 282–283
 cleaning, 282
 drying, 282–283
 sample selection, 292–294
 chemical composition in, 293
 conventional methods, 292
 spectral sample selection, 293–294
 drawbacks, 293
 uniform sample selection, 293
 sample storage, 287–289
 factors affecting, 288
 moisture content, 288
 sample stream splitter, 276
 sampling, 268–272
 sampling methods, types, 271
 automatic, 271
 manual, 271
 no sampling at all, 271
 screw conveyer-type, 748
 stratification, 281
 subsampling, 281–282
 very large objects sampling, 280–281
 woodside sampler, 278
 Saturated aliphatic polyesters, 552
 Savitzsky–Golay method, *see* Polynomial smoothing, in
 Fourier transform
 SCADA (Supervisory Control and Data Acquisition),
 246
 Scattering technique, 210
 of Raman spectroscopy, 13–15
 scattering order, 26
 Schuster's theory, 27–28
 SCMWPLS (searching combination moving window
 partial least squares), 679–681
 Scouring, 468, 469, 470, 472, 477
 SEA, 374
 SEC (size exclusion chromatography), 536, 548

SEC (standard error calibration), 111, 126, 523, 541, 548

SEE (standard error of estimate), 536–562, 592–602

Semisolids, sampling of, 280

Sensitivity and detection limits, NIR, 126

Sensors, *see also* Process sensors

- hard sensors, 247
- in pharmaceutical manufacturing unit, 248
- sensor element, 247
- sensor system, 247
- soft or intelligent sensors, 247

SEP (standard error of performance), 110–111, 536–549, 592–594, 601–602

Shewhart's control chart, 252

Sildenafil citrate, 631

- NIR spectrum of, 634–644

Silicones, 561–562

SIMCA (soft independent modeling of class analogies), 185, 603

Skim milk powder content, NIR analysis of, 426

Slit distribution function, 351

SMLR (stepwise multiple linear regression), 193, 594, 785

Sodium lauryl sulfate, 550

Soft-modeling calibration, 215

SOPs (standard operating procedures), 614

Source, in FT-NIR, 81

Spanish food products, labeling of, 387–390

- Iberian ham, 388
- olive oil, 388–389

SPC (statistical process control), 251–252

- Neave and Wheeler's Shewhart's charts and probability approach in, 252–253

Spectral reconstruction, 333–344

- correlation plots, 335–339
- difference spectra, 339–341
- Honigs reconstruction, 336
- sharpened spectra, 341–344

SRMs (standard reference materials), 522

Standardization methods, in multivariate calibration models, 234–240

- based on correcting predicted values, 237
- based on transferring calibration models, 239
- based on transferring NIR spectra, 237–239
- computation of parameters, 237–240
- approaches to, 237
- principles, 234
- samples

 - number of, 236
 - selection, 234–236
 - approaches for, 234–235
 - selection of the most suitable approach, 235–236
 - stability and representativity of, 234, 236
 - validation of the standardization parameters, 239

Standards, in FT-NIR, 88–91

Statistical terms, in NIR calibration, 140–141; *see also under* Calibration basics

Stokes equations, 43

- for an assembly of sheets, 44

Studentized leverage corrected residuals, 219

Suessen method, 501

Sugars spectra, 624

Superba process, 501

Synchrotron infrared light, 664

Synthetic fibers, yarns, and textile products, 497–519

- at-line/on-line analyses, 512–519
- finish-on-fiber measurements, 499–501
- heatset temperature measurements, 501–503
- heatset temperature prediction, 503–505
- moisture content in nylon fibers, 498–499
- process control measurements, nylon carpet yarns, 505–506
- textile product identification, 506–511

 - carpet fiber identification for recycling, 508–511
 - fiber/polymer-type identification, 507–508
 - nylon and heatsetting type identification, 507
 - polyester fiber by different producers, 507

T

TAME (*tert*-amyl methyl ether), 522

Taping, 289, 572

Target function, 168

Taylor series, 106–107

Test statistics, in NIR calibration, 141–149; *see also under* Calibration basics

Textiles, NIR analysis of, 485–520

- natural fibers, 486–497
- synthetic fibers, yarns, and textile products, 497–519

Theory

- and model systems, in diffuse reflection, 48–59
- carbazole in a matrix of varying absorption, 52–56
- graphite in NaCl, 50–52
- mixture of wheat and rape seed meal, 56–59
- for sheets and an assembly thereof, 42–45

“Thermometrical spectrum,” 3

Thin-cap fibroatheroma, 664

Thinylloestradiol, 588

Throughput advantage (Jacquinot advantage), of FT-NIR, 84

Tilting filter concept, 72–73

TPH (total petroleum hydrocarbon), 526

Transflectance liquid cell, 738

Transflectance, 279

Transmission, in light and matter interaction, 23–24

Transmittance techniques, 69–70

Turquoise, detection of, 768–772

- background, 768–769
- experimentation, 769–772

Turret-mounted discrete filter instrument, 73

Two-flux treatments, 26–27

U

Unsaturated polyesters, 552

V

Variance scaling, 208–210

Viagras authenticity, 631–644

- materials and methods, 632–633
- chemical analyses, 633
- chemometrics, 632–633
- libraries, 632

Viagras authenticity (*continued*)
 NIRS analyses, 632
 samples, 632
 results and discussion, 633–644
 principal component analyses, 635–644
 wavelength correlation, 633–635
Vibrational spectroscopy
 basic principles of, 9–18
 absorption techniques of, 9–13; *see also*
 Absorption techniques
 fluorescence and scattering efficiency in, 14
Vitamins spectra, 624
VLT (variable layer thickness) method, 59–60

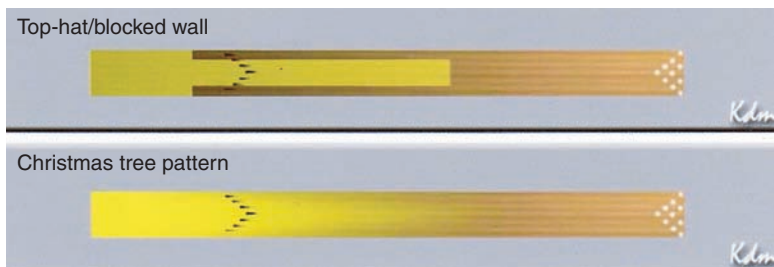
W

Water and hydrogen bonding, 355–360
Wavelet Transformation, 784–785

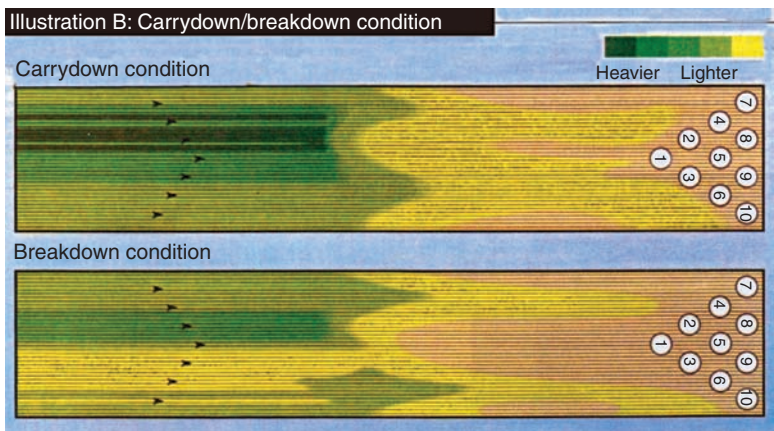
Wedge interference filters, 72
Weighted means, 269
Wines and distilled liquors, 458–459
Wool, NIR analysis of, 465–478
 calibration mathematics, 472–477
 calibration set assembly, 470
 extensions, 477–478
 Gaussian distribution of samples,
 473
 instrumentation, 468–470
 moisture content measurements, 472
 reference methods, 470–472
 yield tests, 466
WRONZ, 467, 470, 473, 477

Y

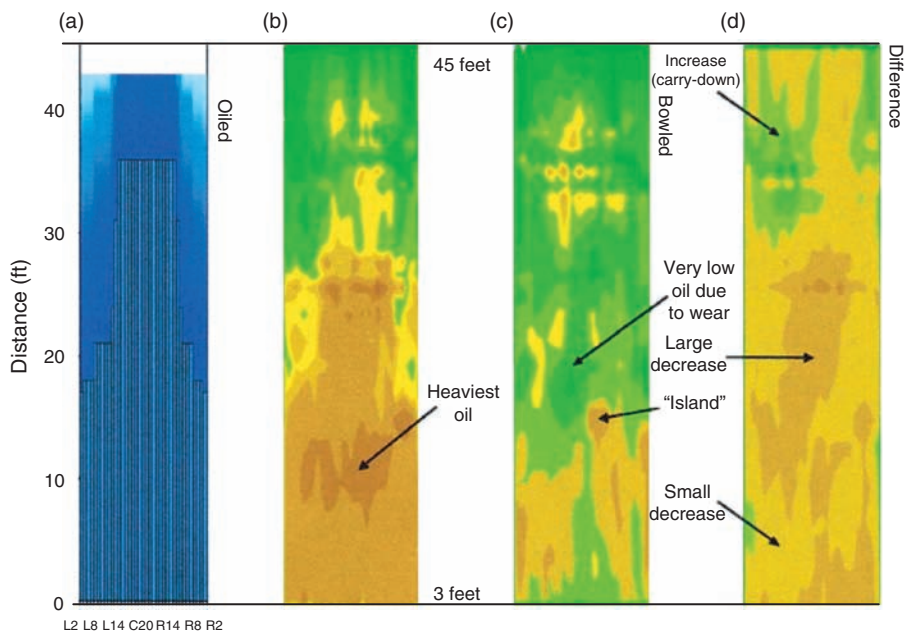
Yield tests, 466, 467



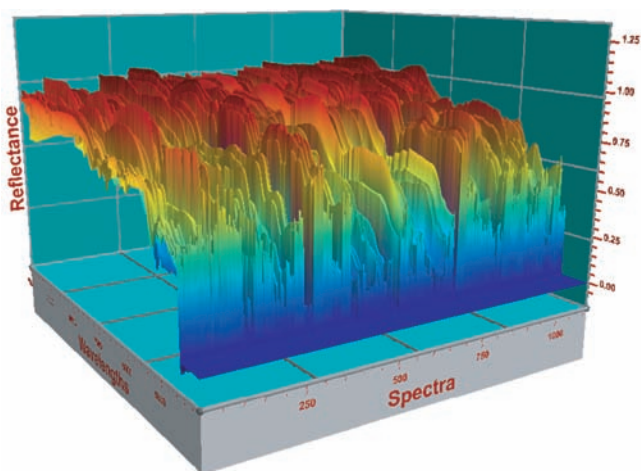
COLOR FIGURE 28.2 Examples of oil patterns.



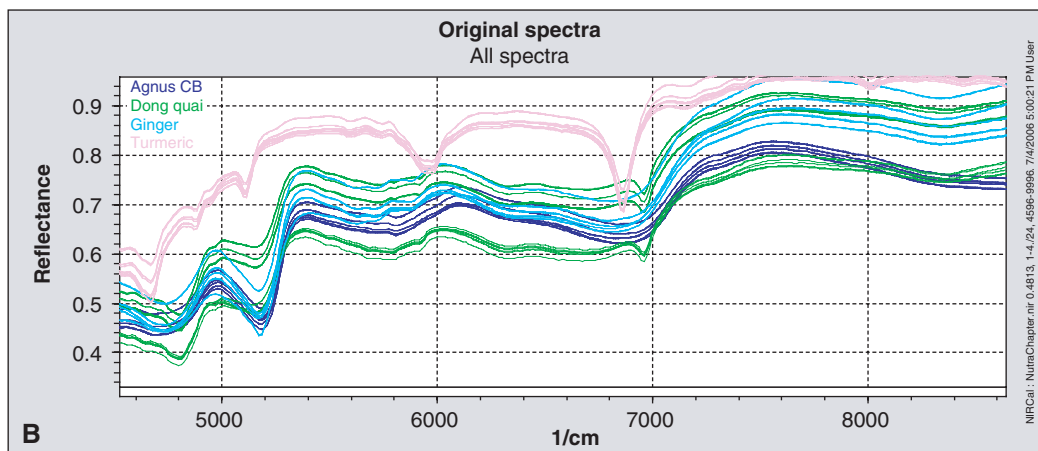
COLOR FIGURE 28.3 Illustrative of carry-down and breakdown oil conditions.



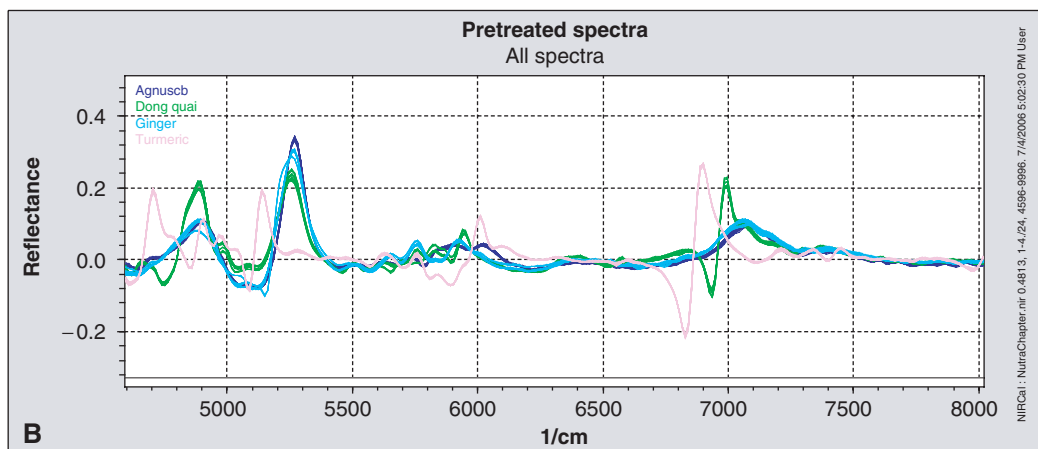
COLOR FIGURE 28.10 (a) Theoretical pattern programmed into the lane oiling machine, (b) NIR-measured lane before bowling, (c) NIR-measured lane after four matches, and (d) before and after difference in NIR maps.



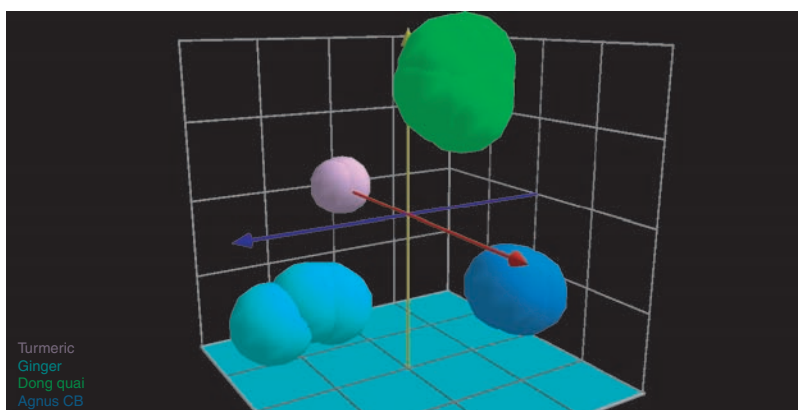
COLOR FIGURE 31.1 Three-dimensional plot of 5 different nutraceutical ingredients with about 25 spectra of each.



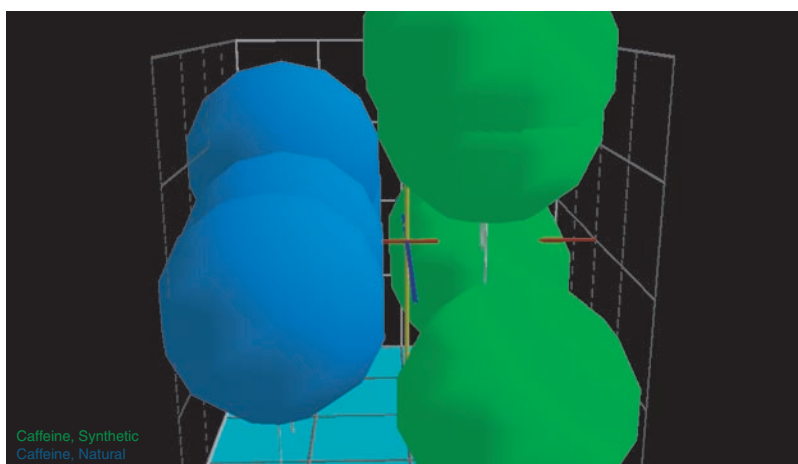
COLOR FIGURE 31.2 Raw spectra for multiple lots of agnus castus berry, dong quai, ginger, and turmeric.



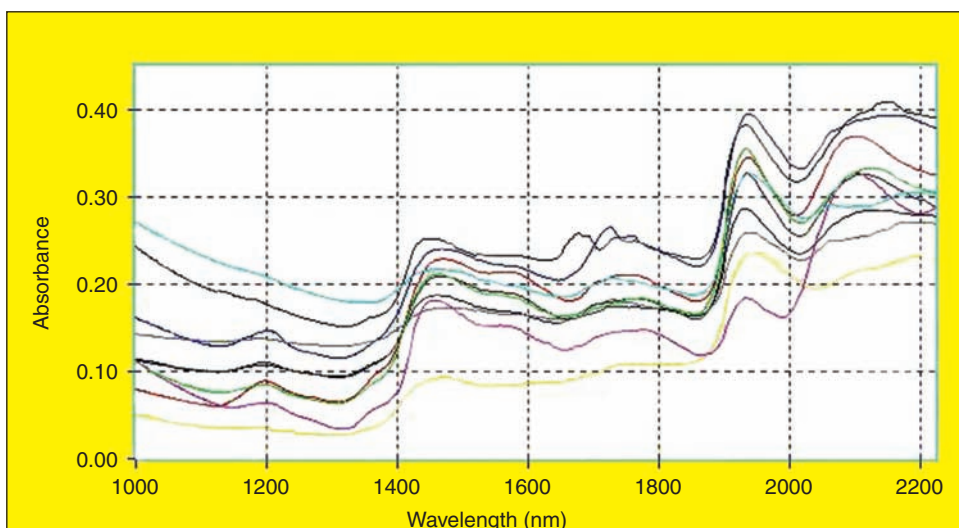
COLOR FIGURE 31.3 First derivative spectra for multiple lots of agnus castus berry, dong quai, ginger, and turmeric.



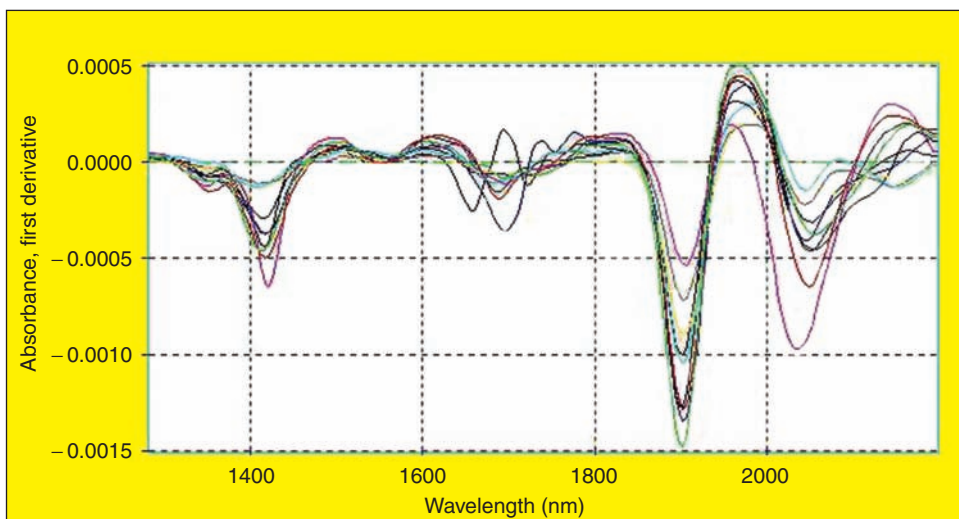
COLOR FIGURE 31.4 Clusters for four different materials in three-dimensional space, clearly showing that the groups can be separated.



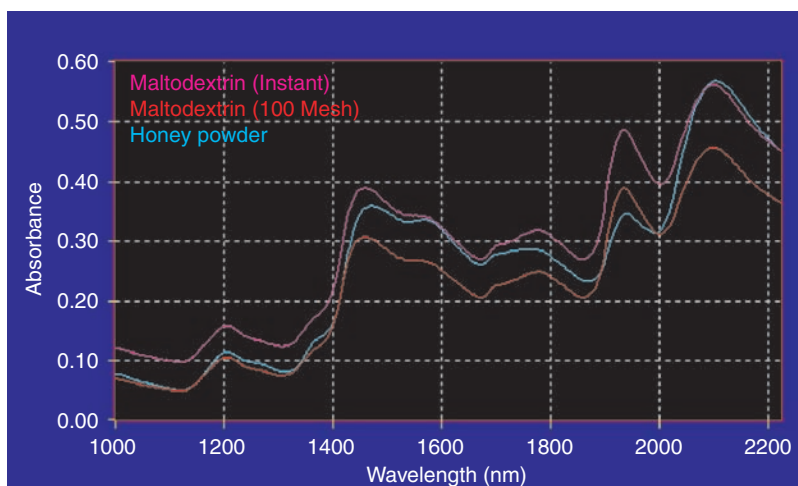
COLOR FIGURE 31.6 Clusters of natural caffeine and synthetic caffeine.



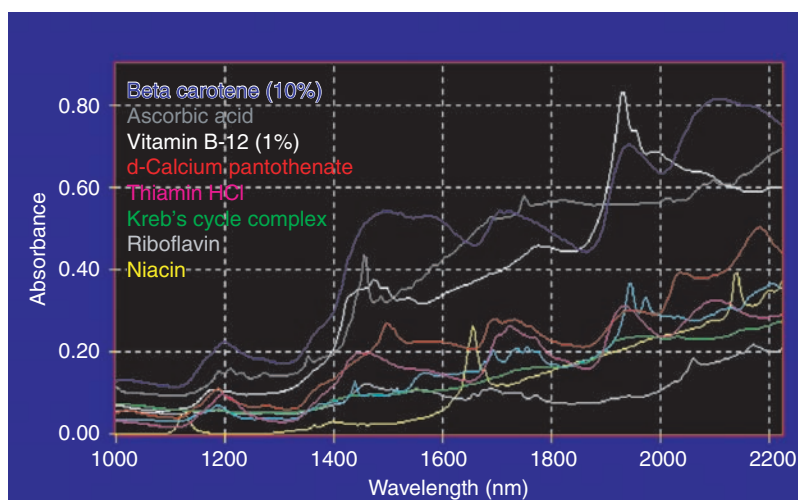
COLOR FIGURE 31.9 Spectra of various herbal materials.



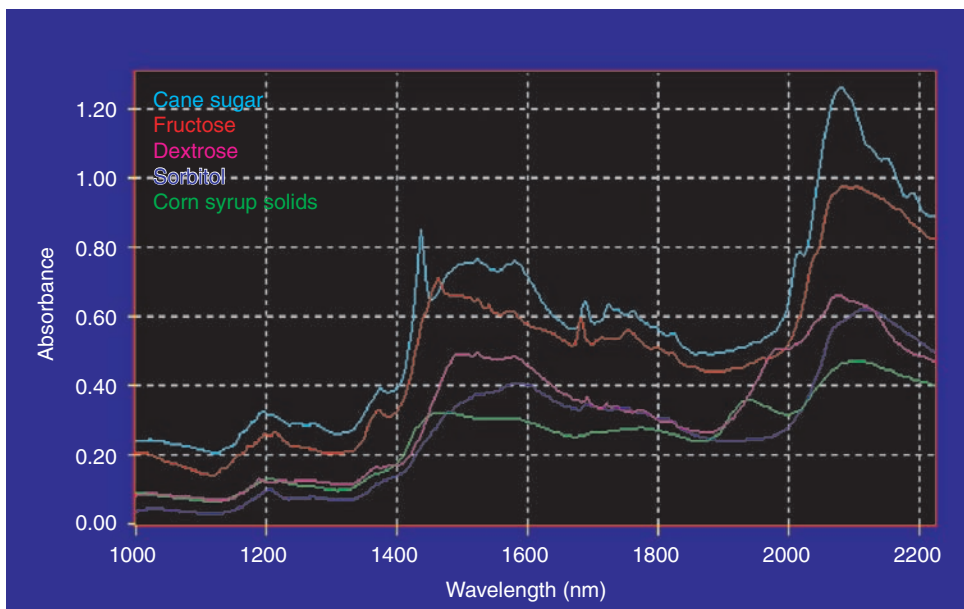
COLOR FIGURE 31.10 First-derivative spectra of various herbal materials.



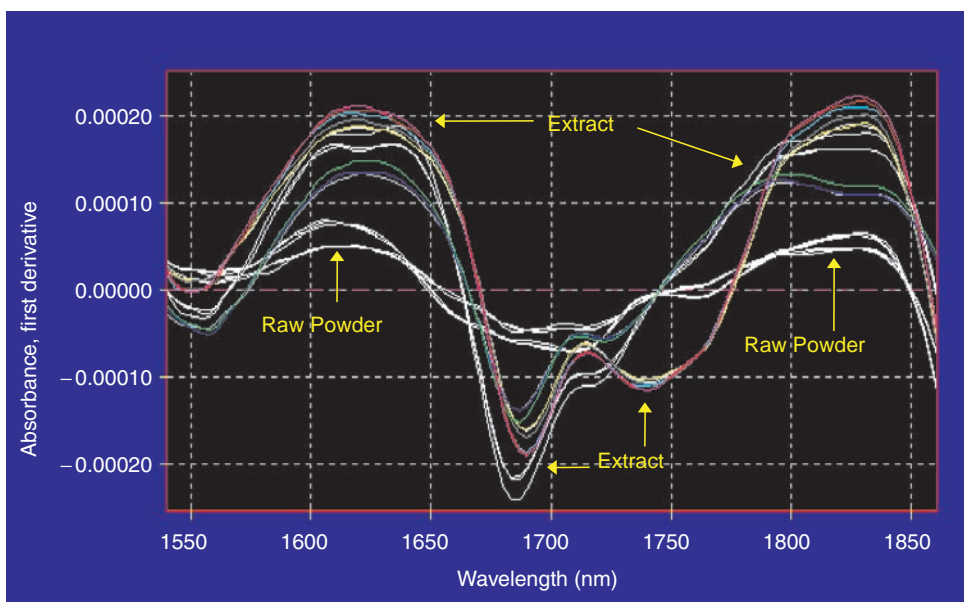
COLOR FIGURE 31.12 Original spectra of various carbohydrates.



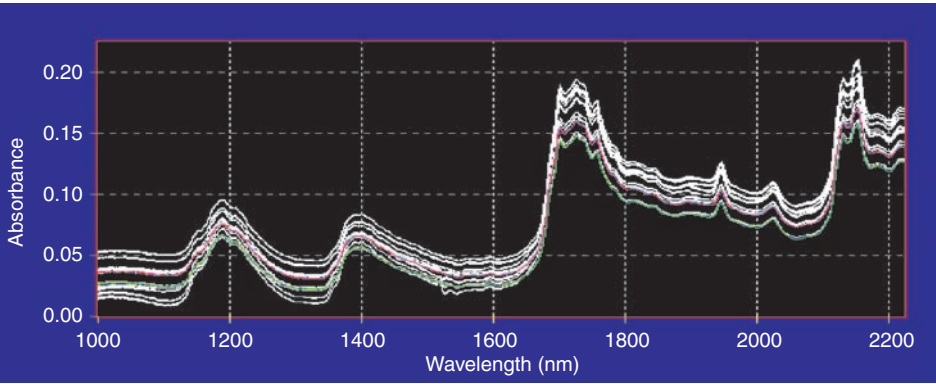
COLOR FIGURE 31.13 Original spectra of various vitamins.



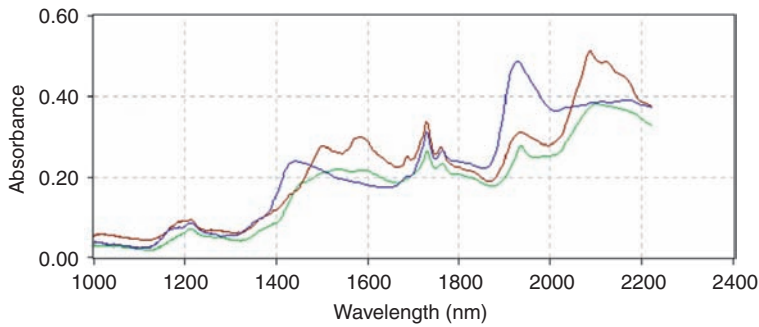
COLOR FIGURE 31.14 Original spectra of various sugars.



COLOR FIGURE 31.15 First-derivative spectra of raw burdock root and burdock root extract.

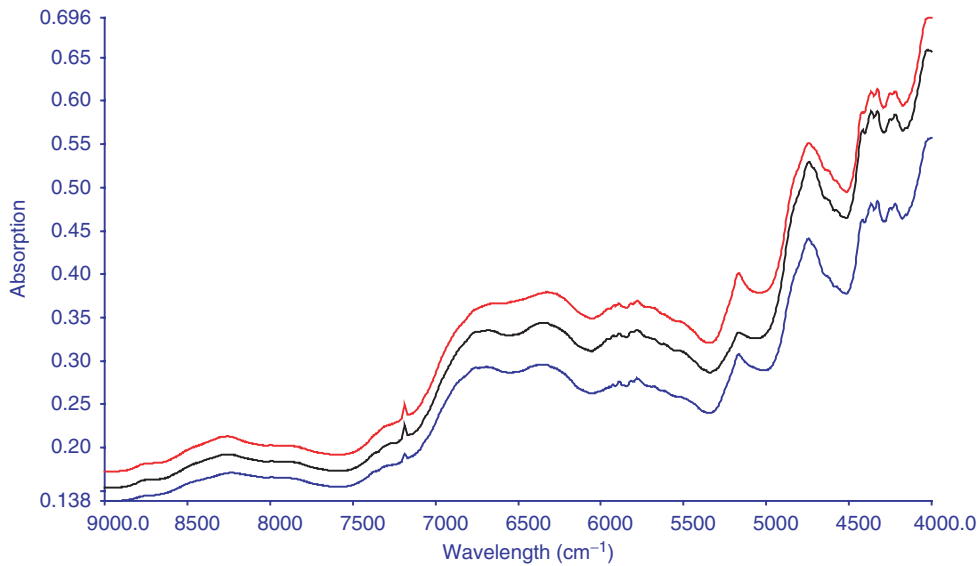


COLOR FIGURE 31.19 Original spectra of 19-Norandrosteine Assay.

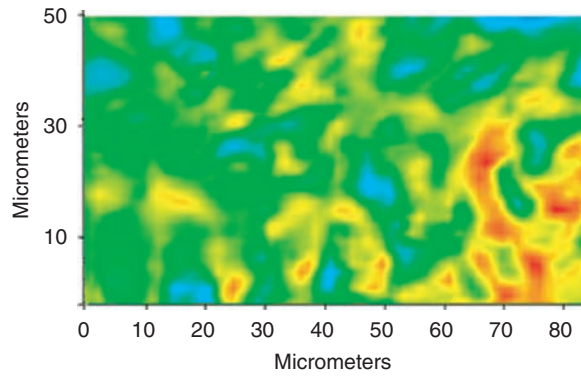


Sample	HPLC	CE	FT-NIR	Carbazole
Red	91.5%	101.6%	<i>Fail</i>	14.6%
Green	96.1%	N/A	<i>Fail</i>	11.4%
Blue	95.8%	99.7%	<i>Pass</i>	99.7%

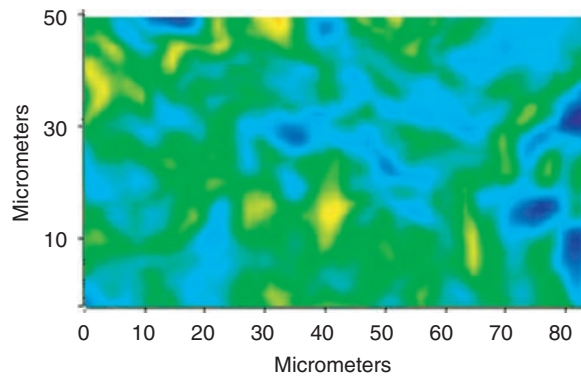
COLOR FIGURE 31.21 FT-NIR detects materials that give false positive by other techniques for chondroitin sulfate.



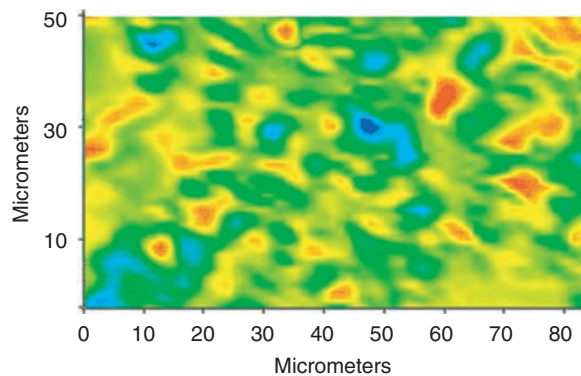
COLOR FIGURE 32.4 NIR spectra of counterfeit tablets from three different batches containing amphetamine as active ingredient.



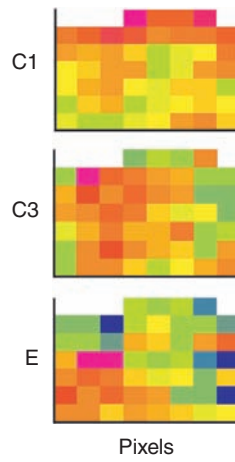
COLOR FIGURE 34.12 Collagen I distribution in targeted region of fibrous cap (red = high, blue = low concentration).



COLOR FIGURE 34.13 Collagen III distribution in targeted region of fibrous cap (yellow = high, blue = low concentration).



COLOR FIGURE 34.14 Elastin distribution in targeted region of fibrous cap (red = high, blue = low concentration).



COLOR FIGURE 34.15 Localized distribution of collagen I, III, and elastin at individual pixel resolution.

Fast, inexpensive, and easy-to-use, near-infrared (NIR) spectroscopy can be used to analyze small samples of virtually any composition. The **Handbook of Near-Infrared Analysis, Third Edition** explains how to perform accurate as well as time- and cost-effective analyses across a growing spectrum of disciplines.

Presenting nearly 50% new and revised material, this thoroughly updated edition incorporates the latest advances in instrumentation, computerization, calibration, and method development in NIR spectroscopy. The book underscores current trends in sample preparation, calibration transfer, process control, data analysis, and commercial NIR instrumentation. New chapters highlight novel applications including the analysis of agro-forestry products, polymers, blood, and control serum. They also cover NIR spectra, process analytical technologies (PAT), quantitative and qualitative analyses for nutraceuticals, NIR photography uses in medicine, and counterfeit detection methods for various applications.

Features

- Provides new chapters on diffuse reflection theories, the calibration and validation of process sensors, and the NIR spectra of gases
- Reviews software-enabled chemometric methods and other trends in data analysis
- Covers on-line, in-line, and at-line analyses, multilinear regression, principal component analysis, and more
- Highlights novel applications in pharmaceuticals, polymers, plastics, petrochemicals, textiles, foods and beverages, baked products, agricultural products, biomedicine, nutraceuticals, and counterfeiting
- Discusses non-traditional applications such as analyzing bowling lane conditions and counterfeit detection for currency, turquoise, and cigars
- Includes new color images of NIR spectra not found in previous editions in addition to more than 400 black and white illustrations and photos

Offering the most complete single-source guide of its kind, the **Handbook of Near-Infrared Analysis, Third Edition** continues to offer practicing chemists and spectroscopists an unparalleled combination of theoretical foundations, cutting-edge applications, and practical experience provided firsthand by more than 60 experts in the field.

7393



CRC Press

Taylor & Francis Group
an **informa** business

www.taylorandfrancisgroup.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

ISBN 0-8493-7393-X



www.crcpress.com